GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 19, 2003, 15:54:33 ; Search time 293 Seconds
(without alignments)
4904.289 Million cell updates/sec

Title: US-09-315-355A-47
Perfect score: 1853
Sequence: 1 gatccctgagcgtgtggcag.....aaaagtaattccttaaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	150	8.1	150	10	US-09-880-107-2725	Sequence 2725, Ap
2	53.2	2.9	2895	9	US-09-938-842A-2540	Sequence 2540, Ap
c 3	52.8	2.8	766	10	US-09-864-761-19608	Sequence 19608, A
c 4	52.8	2.8	1944	10	US-09-864-761-2825	Sequence 2825, Ap
5	49	2.6	1959	10	US-09-864-761-4012	Sequence 4012, Ap
6	47.8	2.6	583	10	US-09-864-761-20772	Sequence 20772, A
7	47	2.5	1626	10	US-09-764-848-19	Sequence 19, Appl
c 8	46.2	2.5	58985	9	US-09-901-152-3	Sequence 3, Appli
c 9	44.6	2.4	439	10	US-09-864-761-20174	Sequence 20174, A
10	44.2	2.4	2142	10	US-09-917-800A-1567	Sequence 1567, Ap
11	43.8	2.4	888	10	US-09-770-445-502	Sequence 502, App
c 12	43.6	2.4	327	10	US-09-864-761-28059	Sequence 28059, A
13	43.6	2.4	575	10	US-09-864-761-20733	Sequence 20733, A
14	43.6	2.4	1969	10	US-09-864-761-3972	Sequence 3972, Ap
15	43.4	2.3	720	9	US-10-101-487-74	Sequence 74, Appl
c 16	43.4	2.3	720	9	US-10-101-487-76	Sequence 76, Appl
c 17	43	2.3	8895	10	US-09-764-853-887	Sequence 887, App
c 18	43	2.3	8895	10	US-09-764-853-937	Sequence 937, App
c 19	43	2.3	9656	10	US-09-764-853-886	Sequence 886, App
c 20	43	2.3	9656	10	US-09-764-853-933	Sequence 933, App
21	42.8	2.3	522	9	US-10-101-487-71	Sequence 71, Appl
c 22	42.8	2.3	530	9	US-10-101-487-73	Sequence 73, Appl
23	42.8	2.3	554	9	US-10-101-487-69	Sequence 69, Appl
24	42.8	2.3	554	9	US-10-101-487-106	Sequence 106, App
c 25	42.4	2.3	565	10	US-09-864-761-6633	Sequence 6633, Ap
26	42.2	2.3	1920	9	US-09-906-393A-35	Sequence 35, Appl
27	42.2	2.3	1958	10	US-09-791-406-3	Sequence 3, Appli
28	42.2	2.3	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c 29	41.4	2.2	460	10	US-09-864-761-2587	Sequence 2587, Ap
30	41.4	2.2	466	10	US-09-864-761-98	Sequence 98, Appl
31	41.4	2.2	582	10	US-09-864-761-8188	Sequence 8188, Ap
c 32	41.2	2.2	475	10	US-09-864-761-1361	Sequence 1361, Ap
c 33	41.2	2.2	512	10	US-09-864-761-18121	Sequence 18121, A
34	40.4	2.2	345	10	US-09-822-263-7	Sequence 7, Appli
35	40.2	2.2	520	9	US-10-184-644-332	Sequence 332, App
36	40.2	2.2	632	10	US-09-864-761-20742	Sequence 20742, A
37	40.2	2.2	1983	10	US-09-864-761-3981	Sequence 3981, Ap
38	39.6	2.1	222	10	US-09-864-761-22453	Sequence 22453, A
39	39.6	2.1	484	10	US-09-864-761-5690	Sequence 5690, Ap
40	39.2	2.1	660	10	US-09-864-761-19488	Sequence 19488, A
41	39.2	2.1	962	10	US-09-864-761-2772	Sequence 2772, Ap
42	39.2	2.1	1881	9	US-09-938-842A-2650	Sequence 2650, Ap
c 43	39	2.1	501	10	US-09-864-761-9004	Sequence 9004, Ap
44	38.8	2.1	342	10	US-09-822-263-11	Sequence 11, Appl
45	38.8	2.1	555	10	US-09-822-263-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-880-107-2725/c
; Sequence 2725, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2725
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N92948
US-09-880-107-2725

Query Match 8.1%; Score 150; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 6.6e-33;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    660 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 91

Qy    720 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 31

Qy    780 CTTCACACTCGGAAGTAAACTTTCAAAAAA 809
          ||||||||||||||||||||||||||||||
Db     30 CTTCACACTCGGAAGTAAACTTTCAAAAAA 1
```

```
RESULT 2
US-09-938-842A-2540
; Sequence 2540, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2540
;   LENGTH: 2895
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2540

  Query Match             2.9%;  Score 53.2;  DB 9;  Length 2895;
  Best Local Similarity   48.4%;  Pred. No. 0.00016;
  Matches  182;  Conservative    0;  Mismatches  188;  Indels    6;  Gaps    1;

Qy     460 GATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTG 519
           |||| | |||   | | |  |    | ||    ||  || | ||   ||| ||
Db    1738 GATCCATATCTGAAAGATGTTGATGATGAAGATGATGAAGAGGATATTGATGACACGACA 1797

Qy     520 ATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTA 579
           | |||||  |  |||    | ||  ||||||    |   |||| | ||     || || ||
Db    1798 GTCAAGCCGACCGATTCGGTGATTATTTGTGCTAGGAATGAAGATGATGTTAGCCATCTA 1857

Qy     580 GAGGTGCATGTTTATAATCAAGAAGAAGACTCTT------TTTATGTACACCATGATATA 633
           |||||  || | ||  |  ||       | ||||       | ||||| || ||| | |||
Db    1858 GAGGTTTATTTATACGAGGAATCGTCTGGCTCTCCAAACATGTATGTTCATCATCACATA 1917

Qy     634 CTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGAT 693
           |  |  | | ||  ||  || ||   | ||||||  ||| |  || ||    ||  ||
Db    1918 ATTATACCAGAATTCCCATTGTGTACTGCATGGCTTGATTGTCCTCTTAAAGGAGGGGAA 1977

Qy     694 TCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTT 753
                 |     |  ||    |||    || ||    || ||  | ||||| |||
Db    1978 AAAGGGAATTTTGTAGCTATTGGTTCAAAGGATACACCGACAATCGAGATATGGGATCTT 2037

Qy     754 GATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAG 813
           ||  |   ||||    |   ||     |   ||| ||| | | |      | |    |
Db    2038 GACGTTAGGGACGAGGTGCTACCTTGTATACAACTAGGAGGAATAGAGGAGATGATAGTG 2097

Qy     814 AAAAAGAAAGGAAAGA 829
           |  ||||||   ||||
Db    2098 AGTAAGAAAAAGAAGA 2113


RESULT 3
US-09-864-761-19608/c
; Sequence 19608, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19608
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL008720.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: SWISSPROT HIT: P17164, EVALUE 4.10e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AW844901.1, EVALUE 2.90e+00
US-09-864-761-19608

  Query Match             2.8%;  Score 52.8;  DB 10;  Length 766;
  Best Local Similarity   48.1%;  Pred. No. 9.2e-05;
  Matches  150;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy     320 ATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAG 379
           ||||||||    |   ||||  |        || || ||||     || ||   | ||   |
Db     325 ATGGTGACAATGGTAATGATGGTGATGAGGAGGAGGATGGCGATGATGGTGATGAAGATG 266

Qy     380 ATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTC 439
           || |   ||  || |||| ||||   |||| |||    |||||||    |       |||
Db     265 ATGATGTTGGCGATGAAGATGACGATGATGAAGAGGAGGTTGGTGATGGTGATGATGGTG 206

Qy     440 TTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAAT 499
           |||  |   | || ||| ||| | | ||   | | | |    || ||    ||  |
Db     205 TTAATGGTGATGGTAGTGATGGTGATGAAGATGATGATGTTGGCGATGAAGATGACGATG 146
```

```
Qy    500 ATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTG 559
          |||||  || | |  |     | |     |||||||      | | |||   |   ||
Db    145 ATGAAGAGGAGGTTGGCGATGGTGATGATGGTGATAATGGCGATGGTAGTGATGGTAATG 86

Qy    560 AACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATG 619
          |  |  || ||    |    |  |   ||| | || || | || |  |    | | |||
Db     85 ATGATAACGAGGAAGAGGATGGTGATGATGGTGATGATGAGGAGGTTGGTGATGATGATG 26

Qy    620 TACACCATGATA 631
                ||||
Db     25 ATGGTGATGAAA 14


RESULT 4
US-09-864-761-2825/c
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2825
;   LENGTH: 1944
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL008720.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

  Query Match             2.8%;  Score 52.8;  DB 10;  Length 1944;
  Best Local Similarity   48.1%;  Pred. No. 0.00016;
  Matches  150;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy    320 ATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAG 379
          ||||||||   |   ||||  |       || || ||||     | ||   | ||    |
Db    562 ATGGTGACAATGGTAATGATGGTGATGAGGAGGAGGATGGCGATGATGGTGATGAAGATG 503

Qy    380 ATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTC 439
          || |   ||  || |||| ||||   ||||  |||    |||||||    |       |||
Db    502 ATGATGTTGGCGATGAAGATGACGATGATGAAGAGGAGGTTGGTGATGGTGATGATGGTG 443

Qy    440 TTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAAT 499
          |||  |   | || ||| ||| | | ||   | | | |   |    | ||    ||  |
Db    442 TTAATGGTGATGGTAGTGATGGTGATGAAGATGATGATGTTGGCGATGAAGATGACGATG 383

Qy    500 ATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTG 559
          |||||   || | |  |     | |     ||||||||      | | |||   |   ||
Db    382 ATGAAGAGGAGGTTGGCGATGGTGATGATGGTGATAATGGCGATGGTAGTGATGGTAATG 323

Qy    560 AACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATG 619
          |  |  || ||    |    |  |   ||| | || || | || |  |    | | |||
Db    322 ATGATAACGAGGAAGAGGATGGTGATGATGGTGATGATGAGGAGGTTGGTGATGATGATG 263

Qy    620 TACACCATGATA 631
                ||||
Db    262 ATGGTGATGAAA 251


RESULT 5
US-09-864-761-4012
; Sequence 4012, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
```

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4012
;   LENGTH: 1959
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL022334.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

  Query Match             2.6%;  Score 49;  DB 10;  Length 1959;
  Best Local Similarity   45.6%;  Pred. No. 0.002;
  Matches  172;  Conservative    0;  Mismatches  205;  Indels    0;  Gaps    0;

Qy    317 AGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACT 376
          | ||||||| |   || ||||||       |||||  ||||||    | |||  | |
Db    388 ATGATGGTGGCTATGATGATGACGCTGACGGTGATGGTGATGATGGTGGTGATGATGGTA 447

Qy    377 TAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGG 436
            || |   |||||| |  |||||    |||| |||   |  ||||||   |      ||
Db    448 ATAATGATAGTGATGATGGTGGTGATGGTGATGATGATGATGGTGGTGATGGTGGTAAGG 507

Qy    437 GTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAAC 496
          ||  |   |   | ||   ||||||| | |||  | | | |       | | |   ||
Db    508 GTGGTGATGATAATGGTGATAATGATGATGATGGTGATGATGGGAATGATGGTGATGATG 567

Qy    497 AATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAG 556
          ||||        |||       |        ||   ||  |   |   |||      |
Db    568 GTGATGATGACAGTGATGATGATGATGGTGATGATGGGGATGATGATGACAGTGATGATG 627

Qy    557 CTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTT 616
          |||  | |   |  | ||   || |    || | ||    |||| || |||  |  |
Db    628 GTGATGATGGTGATGGTGATGATGATGATAGTGATGATGGGGAAGATGATGACAGTGATG 687

Qy    617 ATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTG 676
          ||   |  ||||       |     ||   ||  |   ||||  || ||      ||  ||
Db    688 GTGGTGATGATGACGATAGCGATGATGGGGATGATGACAGTGATGATTGTGATGATGGTG 747

Qy    677 ATCCTAGCCCAGATGAT 693
          ||  |     | ||||
Db    748 ATGATGATGATGGTGAT 764


RESULT 6
US-09-864-761-20772
; Sequence 20772, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20772
;   LENGTH: 583
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL022334.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;   OTHER INFORMATION: EST_HUMAN HIT: AW992395.1, EVALUE 9.50e-02
US-09-864-761-20772

  Query Match             2.6%;  Score 47.8;  DB 10;  Length 583;
  Best Local Similarity   47.0%;  Pred. No. 0.0021;
  Matches  148;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy    316 GAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGAC 375
          || ||||        || |||||  |       ||  |||||||| |     |||  | |
Db     65 GATGATGACAGTGATGATGATGATGGTGATGATGGGGATGATGATGACAGTGATGATGGT 124

Qy    376 TTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTG 435
             |||    ||| ||| || | |||    |||| || |     || |||   |
```

```
Db    125 GATGATGGTGATGGTGATGATGATGATAGTGATGATGGGGAAGATGATGACAGTGATGGT 184

Qy    436 GGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAA 495
Db    185 GGTGATGATGACGATAGCGATGATGGGGATGATGACAGTGATGATTGTGATGATGGTGAT 244

Qy    496 CAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGA 555
Db    245 GATGATGATGGTGATGATGGGGATGATGATCACAGTGATGATGGTGATGGTGATGATGAT 304

Qy    556 GCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTT 615
Db    305 GATGGTGATGACGGGGATGATGATGAGAGTGATGATGGTGATGATGGTGATGATGATGGT 364

Qy    616 TATGTACACCATGAT 630
Db    365 GATGATGACGGTGAT 379


RESULT 7
US-09-764-848-19
; Sequence 19, Application US/09764848
; Patent No. US20020077270A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ08
;  CURRENT APPLICATION NUMBER: US/09/764,848
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 1626
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-848-19

  Query Match             2.5%;  Score 47;  DB 10;  Length 1626;
  Best Local Similarity   51.7%;  Pred. No. 0.0066;
  Matches  104;  Conservative    1;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    208 GAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAA 267
Db    372 GAATATATAGAAAGAGAAGAATCTGATGGTGAATATGATGAGTTTGGACGTAAAAAGAAA 431

Qy    268 GATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGAC 327
Db    432 AAATACAGAGGGAAAGCAGTTGGTCCTGCATCTATATTAAAGGAAGTTGAAGATAAAGAA 491

Qy    328 CCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATAT 387
Db    492 TCAGAGGGAGAAGAAGAGGATGAGGATGAAGATCTTTCTAAATATAAGTYAGATGAGGAT 551

Qy    388 GATGAGGAAGGTGACCCAGAT 408
Db    552 GAGGATGAAGATGACGCTGAT 572


RESULT 8
US-09-901-152-3/c
; Sequence 3, Application US/09901152
; Publication No. US20030022824A1
; GENERAL INFORMATION:
;  APPLICANT: HU, Song et al.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001248
;  CURRENT APPLICATION NUMBER: US/09/901,152
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 58985
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(58985)
;   OTHER INFORMATION: n = A,T,C or G
US-09-901-152-3

  Query Match             2.5%;  Score 46.2;  DB 9;  Length 58985;
  Best Local Similarity   50.7%;  Pred. No. 0.11;
  Matches  111;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy    319 GATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTA 378
Db  11221 GATGATGATGGTGATGATGATGATGGTGATGATGATGGTGATGATGATAATGATGATGAT 11162

Qy    379 GATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGT 438
Db  11161 GGTGGTGATGATGGTGATGGTGATGATGGTGATGGTGATGTTGGTGGTGATGATAATGAT 11102

Qy    439 CTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAA 498
Db  11101 GATGGTGATGATGGTGGTGATGATGATGGTGATGAAGATGATAATGATGATGGTGATGAT 11042

Qy    499 TATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAAT 537
Db  11041 GATGATGGTGATGATGATGATAATGATGATGGTGATAAT 11003


RESULT 9
US-09-864-761-20174/c
; Sequence 20174, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
```

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20174
;   LENGTH: 439
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL035419.9
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-20174

  Query Match             2.4%;  Score 44.6;  DB 10;  Length 439;
  Best Local Similarity   48.6%;  Pred. No. 0.014;
  Matches  122;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    316 GAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGAC 375
          || | || |||    | |||| |    |  ||||| ||||||    | ||| |||||
Db    294 GATGCTGCTGATGGTGTTGATGGCGATGCTGCTGATGGTGATGATGGTGGTGATTACGAT 235

Qy    376 TTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTG 435
            |||     || ||  |  ||||      ||||||   | ||| ||  | |
Db    234 GGTGATGCTGCTGCTGCTGCTGGTGGTGATCCTGCTGATGGTTTTGATGGTTATGATGAT 175

Qy    436 GGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAA 495
          |||  | | |   ||||  | ||| | | |||  | | | || |    | |||   ||
Db    174 GGTGATGCTGCTGACGGTGATGATGGTGATGATGATGATGGTAATGATGATGATGGTGAG 115

Qy    496 CAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGA 555
            || |  || | |||          |     |||||||     | |  ||   |
Db    114 GGTAATAATGGTAATGATGGTGATGGCGATGATGGTGATAATAACGATGATGCTGATGGT 55

Qy    556 GCTGAACAGGA 566
          | |||  | ||
Db     54 GATGATAATGA 44


RESULT 10
US-09-917-800A-1567
; Sequence 1567, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1567
;   LENGTH: 2142
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_012749
US-09-917-800A-1567

  Query Match             2.4%;  Score 44.2;  DB 10;  Length 2142;
  Best Local Similarity   47.9%;  Pred. No. 0.05;
  Matches  127;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy    149 CACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGG 208
          | || | |   | ||   |  | | |||| |   |         || ||| |   | |
Db    551 CTCCTGCCTCAGAGGATGAGGATGAGGAAGATGATGATGATGAAGATGATGATGATGATG 610

Qy    209 AGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAG 268
          |  ||  | | || || |  | || |  ||| ||||| |  |  ||  |  |  |||
Db    611 ATGAAGAGGAGGAGGAGGAAGATGACTCTGAGGAAGAAGTTATGGAGATCACACCAGCCA 670

Qy    269 ATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACC 328
          | || | | | | | |    |      |  |      | ||       | ||  |  ||
Db    671 AAGGAAAGAAAACTCCTGCAAAAGTTGTTCCTGTGAAAGCCAAGAGTGTGGCCGAGGAGG 730

Qy    329 CAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATG 388
            || |||||    |      ||| ||||| |||  || |||  | |||   |||| || | |
Db    731 AGGAAGATGATGAGGATGATGAAGATGAAGAGGAGGATGAAGATGAAGAAGATGAAGAGG 790

Qy    389 ATGAGGAAGGTGACCCAGATGCTGA 413
          | || |||| |||    || |  ||
Db    791 ACGATGAAGATGAGGATGAGGAAGA 815


RESULT 11
US-09-770-445-502
; Sequence 502, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
;  APPLICANT:  Gorlach, Jorn
;  APPLICANT:  An, Yong-Qiang
;  APPLICANT:  Hamilton, Carol M.
;  APPLICANT:  Price, Jennifer L.
;  APPLICANT:  Raines, Tracy M.
;  APPLICANT:  Yu, Yang
;  APPLICANT:  Rameaka, Joshua G.
;  APPLICANT:  Page, Amy
;  APPLICANT:  Matthew, Abraham V.
;  APPLICANT:  Ledford, Brooke L.
;  APPLICANT:  Woessner, Jeffrey P.
;  APPLICANT:  Haas, William David
;  APPLICANT:  Garcia, Carlos A.
;  APPLICANT:  Kricker, Maja
;  APPLICANT:  Slader, Ted
;  APPLICANT:  Davis, Keith R.
;  APPLICANT:  Allen, Keith
;  APPLICANT:  Hoffman, Neil
;  APPLICANT:  Hurban, Patrick
;  TITLE OF INVENTION:  Expressed Sequences of Arabidopsis
```

```
;  TITLE OF INVENTION:  thaliana
;  FILE REFERENCE:  2023US (PARA-012PRV)
;  CURRENT APPLICATION NUMBER: US/09/770,445
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER:  US 60/178,472
;  PRIOR FILING DATE:  2000-01-27
;  NUMBER OF SEQ ID NOS:  999
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 502
;   LENGTH: 888
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(888)
;   OTHER INFORMATION: n = A,T,C or G
US-09-770-445-502

  Query Match             2.4%;  Score 43.8;  DB 10;  Length 888;
  Best Local Similarity   46.8%;  Pred. No. 0.037;
  Matches  138;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy    131 GCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCC 190
          ||| ||  |  |  ||   |  | |       |||  |  || ||||||      |
Db     35 GCGACGACGATAGTGATGAAGAACAAGTTAAAGACGTAGGTGAAGAAGTTAGCGACGAAG 94

Qy    191 TCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGA 250
            |  || ||| | ||  ||    |    | || || | ||  |     || ||||| ||
Db     95 ACGATGATGATGGAAGTGAAGGAGATGATGATGATGATGAGGAAGAAGAAGAAGATGACG 154

Qy    251 CAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGC 310
            | |  |  |   |   | |   |||||  |  | |     |   |     | |  || |
Db    155 ATGACGATGATGATGTTCAGGTTTTGCAGTCTTTAGGTGGTCCTCCTGTACAATCAGCGG 214

Qy    311 CCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGT 370
              || ||||  ||    || || ||  |||    |||||| |||||    | ||
Db    215 AAGATGAAGATGAAGAAGGTGATGAAGATGGGAACGGTGATGACGATGATGATGATGGAG 274

Qy    371 ACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGA 425
          | ||    ||| |  | ||||| || |  ||    |||| |||   |   || ||
Db    275 ATGATGATGATGACGACGATGATGATGAGGATGAGGATGTTGAAGATGAGGGGGA 329


RESULT 12
US-09-864-761-28059/c
; Sequence 28059, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING.DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28059
;   LENGTH: 327
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005668.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-28059

  Query Match             2.4%;  Score 43.6;  DB 10;  Length 327;
  Best Local Similarity   47.2%;  Pred. No. 0.023;
  Matches  133;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    321 TGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGA 380
          || |||    || ||||||||  |    |||||||||||    |  |||  | ||    ||
Db    306 TGATGATGATGATGATGACAGCAGTGATGATGATGATGATGAGAATGATGATGATGATGA 247

Qy    381 TAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCT 440
            |  ||||||| ||  | ||     ||| |||      || |||  .|     | |
Db    246 CAGTGATGATGATGACAGCGATGACAATGATGATGACAGTGATGATGATGATGATGATGA 187

Qy    441 TACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATA 500
          |  |   | |     | |||| |   |  | | | |        | | |   ||  | |
Db    186 CAGTGATGATGATGATGATGATGACAGTGATGATGATGACAGTGATGGTGATGATGATGA 127

Qy    501 TGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGA 560
          |||  |||| | |       | |      |||| ||  |   |||     ||   | |||
Db    126 TGACAGTGATGGTGATGATGATGACAGTGATGATGATGATGATGTTGACAGCGATGGTGA 67

Qy    561 ACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGA 602
            | ||  |  || ||   || |     | | |  || | ||
Db     66 TGATGATGACAGCGATGATGATGACAGCGATGACGATGATGA 25


RESULT 13
US-09-864-761-20733
; Sequence 20733, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
```

```
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20733
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  12
;   OTHER INFORMATION: NT HIT: AF185255.1, EVALUE 4.80e+00
;   OTHER INFORMATION: EST_HUMAN HIT: BE071749.1, EVALUE 5.80e+00
US-09-864-761-20733

  Query Match             2.4%;  Score 43.6;  DB 10;  Length 575;
  Best Local Similarity   47.7%;  Pred. No. 0.032;
  Matches  127;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    153 AGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAA 212
          |||  |||  |||       | || ||||  ||          |  |||| | |||||||
Db    281 AGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGAAGAGGAGAA 340

Qy    213 ATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGG 272
          | |||| || |  |  |  |  || || ||||||  ||  |          || || |
Db    341 GAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGA 400

Qy    273 CATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGA 332
          |  |||  | |       | |           ||       | ||  |    ||    ||
Db    401 AGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAAGGAAGAGGA 460

Qy    333 GGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGA 392
          |||  ||    |       || ||  | |||    |  ||| | ||   |||  || | || ||
Db    461 GGAGGAAGAGGAAGAAGAGGAGAAGGAGAAAGAAGAGGAGGAGGAAGAGGAAGAGGAGGA 520

Qy    393 GGAAGGTGACCCAGATGCTGAGACTC 418
          |||||  ||   |  ||  ||||||| |
Db    521 GGAAGCTGGACTTGACACTGAGACAC 546


RESULT 14
US-09-864-761-3972
; Sequence 3972, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3972
;   LENGTH: 1969
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
```

```
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  12
US-09-864-761-3972

  Query Match             2.4%;  Score 43.6;  DB 10;  Length 1969;
  Best Local Similarity   47.7%;  Pred. No. 0.07;
  Matches  127;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    153 AGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAA 212
          |||  |||  |||        | || ||||  ||           |  |||| | |||||||
Db    623 AGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGAAGAGGAGAA 682

Qy    213 ATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGG 272
          | |||| || |  |  |  |  || || ||||||  ||  |         || || |
Db    683 GAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAAGAAGGAAGAGGAGGAGA 742

Qy    273 CATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGA 332
          | |||  | |        | |              ||       | || |    ||    ||
Db    743 AGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAAGGAAGAGGA 802

Qy    333 GGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGA 392
          ||| ||   |       || ||  | |||   |  ||| | ||   |||  || | || ||
Db    803 GGAGGAAGAGGAAGAAGAGGAGAAGGAGAAAGAAGAGGAGGAGGAAGAGGAAGAGGAGGA 862

Qy    393 GGAAGGTGACCCAGATGCTGAGACTC 418
          ||||| ||  |  ||  ||||||| |
Db    863 GGAAGCTGGACTTGACACTGAGACAC 888


RESULT 15
US-10-101-487-74
; Sequence 74, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2)..(718)
US-10-101-487-74

  Query Match             2.3%;  Score 43.4;  DB 9;  Length 720;
  Best Local Similarity   47.9%;  Pred. No. 0.042;
  Matches  125;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy    154 GACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAA 213
          ||  ||| ||||  |    | || ||||   ||         |  |||| | ||||  |
Db    158 GAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAG 217

Qy    214 TTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGC 273
            | |||| || |  |  |     || |||||||||  ||           |||| |
Db    218 GAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAG 277

Qy    274 ATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAG 333
            ||  |          |||          ||||       |||| |  ||   ||||
Db    278 GAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAGGAAGAGGAGGAAGAGGAG 337

Qy    334 GATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAG 393
          || ||   |  |   || || || |||   |  ||| | ||   |||  |  | || |||
Db    338 GAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAG 397

Qy    394 GAAGGTGACCCAGATGCTGAG 414
          ||||  ||    || |  |||
Db    398 GAAGAGGAAGAGGAGGAAGAG 418


Search completed: March 19, 2003, 21:24:48
Job time : 474 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 18, 2003, 19:18:24 ; Search time 60901 Seconds
(without alignments)
885.494 Million cell updates/sec

Title: US-09-315-355A-47
Perfect score: 1853
Sequence: 1 gatccctgagcgtgtggcag.....aaaagtaattccttaaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1853	100.0	1853	9	HUM56KDAPR	L07758 Human IEF S
	2	1833	98.9	1888	9	BC001652	BC001652 Homo sapi
	3	1153.2	62.2	2388	10	BC023137	BC023137 Mus muscu
	4	1152.4	62.2	1828	10	BC003199	BC003199 Mus muscu
c	5	986.6	53.2	164664	2	AC098211	AC098211 Rattus no
	6	784.8	42.4	807	9	BC000067	BC000067 Homo sapi
c	7	373.4	20.2	121700	9	AC007622	AC007622 Homo sapi
c	8	373.4	20.2	334597	2	AC010187	AC010187 Homo sapi
	9	337.6	18.2	400	9	HSXS153	Z36805 H.sapiens (
	10	285	15.4	112300	2	AC109397	AC109397 Rattus no
	11	260	14.0	260	11	G06393	G06393 human STS W
	12	242.8	13.1	250	9	HSXS152	Z36804 H.sapiens (
	13	188.4	10.2	334597	2	AC010187	AC010187 Homo sapi
c	14	150	8.1	150	6	AX410079	AX410079 Sequence
	15	149.8	8.1	38000	8	SPBC1711	AL353012 S.pombe c
c	16	132	7.1	222520	10	AC063968	AC063968 Genomic s
	17	132	7.1	229961	2	AC122919	AC122919 Mus muscu
c	18	129.2	7.0	192874	2	AC112739	AC112739 Rattus no
c	19	124	6.7	40893	3	CBRG45E19	AC084631 Caenorhab
	20	118.2	6.4	1699	9	AB049913	AB049913 Macaca fa
c	21	118	6.4	39553	3	CEJC8	Z82274 Caenorhabdi
	22	108.8	5.9	1558	3	AY075465	AY075465 Drosophil
	23	108.8	5.9	55227	3	AC005437	AC005437 Drosophil
	24	108.8	5.9	140940	3	AC008260	AC008260 Drosophil
	25	108.8	5.9	257909	3	AE003828	AE003828 Drosophil
c	26	108.6	5.9	330724	2	CEY67H2	AL022475 Caenorhab
	27	107.6	5.8	1516	8	AY113942	AY113942 Arabidops
	28	107.6	5.8	1793	8	AY056430	AY056430 Arabidops
	29	107.6	5.8	1806	8	AY050883	AY050883 Arabidops
	30	107.2	5.8	71896	2	AC017793	AC017793 Drosophil
	31	105.8	5.7	184013	2	AC112632	AC112632 Rattus no
c	32	98	5.3	153846	2	AC023252	AC023252 Homo sapi
	33	93	5.0	29379	2	AC084040	AC084040 Homo sapi
	34	84.8	4.6	78220	2	AC023212	AC023212 Homo sapi
	35	82.2	4.4	105470	2	AC116306	AC116306 Dictyoste
	36	74.6	4.0	192874	2	AC112739	AC112739 Rattus no
c	37	69.6	3.8	78220	2	AC023212	AC023212 Homo sapi
c	38	66.6	3.6	115758	9	AC104634	AC104634 Homo sapi
	39	64.4	3.5	61052	2	AC117074	AC117074 Dictyoste
	40	64.4	3.5	61052	2	AC123513	AC123513 Dictyoste
	41	62.8	3.4	222546	2	AC096302	AC096302 Rattus no
	42	62.4	3.4	26085	2	AC096530	AC096530 Rattus no
c	43	62.2	3.4	7218	6	I66494	I66494 Sequence 14
	44	61.6	3.3	186278	9	AC079176	AC079176 Homo sapi
c	45	61.4	3.3	163243	9	AL513304	AL513304 Human DNA

ALIGNMENTS

RESULT 1
HUM56KDAPR
LOCUS HUM56KDAPR 1853 bp mRNA linear PRI 20-APR-1995
DEFINITION Human IEF SSP 9502 mRNA, complete cds.
ACCESSION L07758
VERSION L07758.1 GI:177764
KEYWORDS .
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1853)
AUTHORS Honore,B., Leffers,H., Madsen,P. and Celis,J.E.
TITLE Cloning of a cDNA encoding a novel human nuclear phosphoprotein
belonging to the WD-40 family

```
  JOURNAL   Gene 151 (1-2), 291-296 (1994)
  MEDLINE   95129878
   PUBMED   7828893
FEATURES             Location/Qualifiers
     source          1. .1853
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="MRC-5 V2"
                     /cell_type="fibroblast"
                     /tissue_lib="lambda ZAPII"
     CDS             88. .1593
                     /note="nuclear phosphoprotein (similarity to Saccharomyces
                     cerevisiae PWP1 protein)"
                     /codon_start=1
                     /product="IEF SSP 9502"
                     /protein_id="AAA65201.1"
                     /db_xref="GI:177765"
                     /translation="MNRSRQVTCVAWVRCGVAKETPDKVELSKEEVKRLIAEAKEKLQ
                     EEGGGSDEEETGSPSEDGMQSARTQARPREPLEDGDPEDDRTLDDDELAEYDLDKYDE
                     EGDPDAETLGESLLGLTVYGSNDQDPYVTLKDTEQYEREDFLIKPSDNLIVCGRAEQD
                     QCNLEVHVYNQEEDSFYVHHDILLSAYPLSVEWLNFDPSPDDSTGNYIAVGNMTPVIE
                     VWDLDIVDSLEPVFTLGSKLSKKKKKKGKKSSSAEGHTDAVLDLSWNKLIRNVLASAS
                     ADNTVILWDMSLGKPAASLAVHTDKVQTLQFHPFEAQTLISGSYDKSVALYDCRSPDE
                     SHRMWRFSGQIERVTWNHFSPCHFLASTDDGFVYNLDARSDKPIFTLNAHNDEISGLD
                     LSSQIKGCLVTASADKYVKIWDILGDRPSLVHSRDMKMGVLFCSSCCPDLPFIYAFGG
                     QKEGLRVWDISTVSSVNEAFGRRERLVLGSARNSSISGPFGSRSSDTPMES"
     polyA_signal    1817. .1822
     polyA_site      1853
BASE COUNT      528 a    368 c    461 g    496 t
ORIGIN

  Query Match             100.0%;  Score 1853;  DB 9;  Length 1853;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1853;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 60

Qy      61 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 120

Qy     121 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 180

Qy     181 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 240

Qy     241 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 300

Qy     301 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 360

Qy     361 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 420

Qy     421 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 480

Qy     481 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 540

Qy     541 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     541 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAA 600

Qy     601 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 660

Qy     661 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 720

Qy     721 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 780

Qy     781 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 840

Qy     841 GCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTT 900

Qy     901 TTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCA 960

Qy     961 GCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAA 1020

Qy    1021 GCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGT 1080

Qy    1081 CCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAAT 1140

Qy    1141 CACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGAT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGAT 1200

Qy    1201 GCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260

Qy    1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320

Qy    1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380

Qy    1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440

Qy    1441 AAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 AAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGA 1500

Qy    1501 AGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGC 1560

Qy    1561 AGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTGCTTACC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 AGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTGCTTACC 1620

Qy    1621 TTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAACCATGCA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAACCATGCA 1680
```

```
Qy   1681 GAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGTGGCACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGTGGCACC 1740

Qy   1741 ACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTTGTTGCA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTTGTTGCA 1800

Qy   1801 TTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATTCCTTAAACAT 1853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATTCCTTAAACAT 1853
```

RESULT 2
BC001652

```
LOCUS       BC001652                1888 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, nuclear phosphoprotein similar to S. cerevisiae PWP1,
            clone MGC:2763 IMAGE:2924215, mRNA, complete cds.
ACCESSION   BC001652
VERSION     BC001652.1  GI:12804486
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1888)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: g Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .1888
                     /organism="Homo sapiens"
                     /db_xref="LocusID:11137"
                     /db_xref="taxon:9606"
                     /clone="MGC:2763 IMAGE:2924215"
                     /tissue_type="Kidney, renal cell adenocarcinoma"
                     /clone_lib="NIH_MGC_14"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             97. .1602
                     /codon_start=1
                     /product="nuclear phosphoprotein similar to S. cerevisiae
                     PWP1"
                     /protein_id="AAH01652.1"
                     /db_xref="GI:12804487"
                     /translation="MNRSRQVTCVAWVRCGVAKETPDKVELSKEEVKRLIAEAKEKLQ
                     EEGGGSDEEETGSPSEDGMQSARTQARPREPLEDGDPEDDRTLDDDELAEYDLDKYDE
                     EGDPDAETLGESLLGLTVYGSNDQDPYVTLKDTEQYEREDFLIKPSDNLIVCGRAEQD
                     QCNLDVHVYNQEEDSFYVHHDILLSAYPLSVEWLNFDPSPDDSTGNYIAVGNMTPVIE
                     VWDLDIVDSLEPVFTLGSKLSKKKKKKGKKSSSAEGHTDAVLDLSWNKLIRNVLASAS
                     ADNTVILWDMSLGKPAASLAVHTDKVQTLQFHPFEAQTLISGSYDKSVALYDCRSPDE
                     SHRMWRFSGQIERVTWNHFSPCHFLASTDDGFVYNLDARSDKPIFTLNAHNDEISGLD
                     LSSQIKGCLVTASADKYVKIWDILGDRPSLVHSRDMKMGVLFCSSCCPDLPFIYALRG
                     QKEGLRVWDISTVSSVNEAFGRRERLVLGSARNSSISGPFGSRSSDTPMES"
BASE COUNT      566 a    370 c    463 g    489 t
ORIGIN
```

Query Match 98.9%; Score 1833; DB 9; Length 1888;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1836; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 69

Qy     61 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 129

Qy    121 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 189

Qy    181 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 249

Qy    241 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 309

Qy    301 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 369

Qy    361 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 429

Qy    421 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 489

Qy    481 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 549

Qy    541 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAA 600
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    550 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGACGTGCATGTTTATAATCAA 609

Qy    601 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 669

Qy    661 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 729

Qy    721 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 789

Qy    781 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 849

Qy    841 GCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    850 GCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTT 909

Qy    901 TTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    910 TTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCA 969
```

```
Qy     961 GCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     970 GCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAA 1029

Qy    1021 GCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1030 GCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGT 1089

Qy    1081 CCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090 CCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAAT 1149

Qy    1141 CACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGAT 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1150 CACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGAT 1209

Qy    1201 GCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1210 GCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1269

Qy    1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1270 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1329

Qy    1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1330 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1389

Qy    1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||| || ||||||||
Db    1390 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCCTTCGAGGTCAA 1449

Qy    1441 AAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1450 AAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGA 1509

Qy    1501 AGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGC 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1510 AGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGC 1569

Qy    1561 AGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTGCTTACC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1570 AGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTGCTTACC 1629

Qy    1621 TTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAACCATGCA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1630 TTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAACCATGCA 1689

Qy    1681 GAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGTGGCACC 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1690 GAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGTGGCACC 1749

Qy    1741 ACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTTGTTGCA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1750 ACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTTGTTGCA 1809

Qy    1801 TTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAA 1841
           ||||||||||||||||||||||||||||||||||||  ||
Db    1810 TTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAAAAA 1850
```

```
RESULT 3
BC023137
LOCUS       BC023137                2388 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 2310058A11 gene, clone MGC:28175
            IMAGE:3986908, mRNA, complete cds.
ACCESSION   BC023137
VERSION     BC023137.1  GI:18605706
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2388)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 36 Row: g Column: 9
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .2388
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /map="C57BL/6J"
                      /clone="MGC:28175 IMAGE:3986908"
                      /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                      old, gross tissue."
                      /clone_lib="NCI_CGAP_Mam5"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              63. .1568
                      /codon_start=1
                      /product="RIKEN cDNA 2310058A11 gene"
                      /protein_id="AAH23137.1"
                      /db_xref="GI:18605707"
                      /db_xref="LocusID:67485"
                      /translation="MNRSRQVTCVAWVRRGVAKETPDKVELSKEEVNRLIAEAKGKLQ
                      EEGGSEEEEAGNPSEDGMQSGPTQAPPRESLEDGDPQDDRTLDDDELAGYDLDNYDEE
                      DNPDAETIGESLLGLTVYGSNDQDPYVTLKDTEQYEHEDFLIKPTDNLIVCGRAEQEQ
                      CNLEVHVYNQEEESFYVHHDILLSAYPLSVEWLNFDPSPDASTGNYIAVGNMTPVIEV
                      WDLDIVDSLEPVFTLGSKLSKKKKKKGKKSSSAEGHTDAVLDLSWNKTVRNVLASASA
                      DSTVVLWDLSVGKSVARLTAHTDKVQTLQFHPFEAQTLISGSYDKSVALYDCRDPSQN
                      HRQWRFSGQIERVTWNHFSPCHFLASTDDGFVYNLDARSDKPIFTLNAHNDEISGLDL
                      SSQIKGCLVTASADKFVKIWDILGDRPSLIHSRDMKMGVLFCSSCCPDLPFVYAFGGQ
                      KEGLRVWDISTVSSVNEAFGRRERLVIGGTKGLSVSGPCGSRSPQQTPMES"
BASE COUNT      673 a    526 c    571 g    618 t
ORIGIN
```

```
  Query Match          62.2%;  Score 1153.2;  DB 10;  Length 2388;
  Best Local Similarity  83.6%;  Pred. No. 1.2e-302;
  Matches 1357;  Conservative    0;  Mismatches  258;  Indels    9;  Gaps    4;

Qy      39 TCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAG 98
           ||  |   ||  || ||  || | | ||   ||   |  |   || ||| ||||||||||||
Db      14 TCTCCCAGCTCCCAGGGCTCCAACCTGCGCCGGGTTCAAACCGAAGATCATGAACCGCAG 73

Qy      99 CCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAA 158
           ||||||  |||||  ||||||||||||||||||  ||||||||||||||||  ||||||||
Db      74 CCGCCAAGTGACCTGCGTGGCCTGGGTCCGCCGCGGCGTGGCCAAAGAGACTCCAGACAA 133

Qy     159 GGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCA 218
           |||||||||  ||||||||  |||||  ||  ||  |||||  |||||  |||||  |  ||  |||||
Db     134 GGTAGAGCTAAGTAAAGAGGAAGTGAACCGTCTCATCGCTGAAGCAAAAGGAAAGTTGCA 193
```

```
Qy    219 AGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCA 278
Db    194 AGAAGAA---GGTGGGAGTGAAGAAGAGGAGGCAGGCAACCCTTCAGAAGATGGCATGCA 250

Qy    279 GAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGA 338
Db    251 GAGTGGCCCGACCCAGGCACCGCCTAGGGAGTCCCTGGAGGATGGGGATCCACAAGATGA 310

Qy    339 CAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGG 398
Db    311 CCGCACACTGGACGATGACGAGCTGGCGGGGTACGACTTAGATAATTATGACGAAGAGGA 370

Qy    399 TGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAA 458
Db    371 TAACCCAGATGCTGAGACCATTGGCGAGTCTCTGTTGGGTCTTACGGTCTATGGCAGTAA 430

Qy    459 TGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTT 518
Db    431 TGACCAGGATCCATATGTCACCTTGAAAGACACGGAACAATATGAGCATGAAGATTTCTT 490

Qy    519 GATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTT 578
Db    491 GATTAAGCCGACTGACAACCTCATCGTTTGTGGCCGTGCTGAACAAGAGCAGTGCAATCT 550

Qy    579 AGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTT 638
Db    551 AGAGGTGCACGTTTACAATCAAGAAGAGGAGTCTTTCTACGTCCACCATGACATACTCCT 610

Qy    639 GTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTAC 698
Db    611 GTCTGCATACCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGCATCTAC 670

Qy    699 TGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATAT 758
Db    671 TGGGAATTACATTGCTGTGGGAAACATGACTCCTGTCATCGAGGTATGGGACCTGGATAT 730

Qy    759 AGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAA 818
Db    731 AGTGGACTCTCTAGAACCAGTCTTCACACTGGGAAGCAAGCTTTCCAAAAAGAAGAAGAA 790

Qy    819 GAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATG 878
Db    791 GAAAGGGAAGAAGAGTTCCTCAGCAGAAGGACATACTGATGCTGTCCTTGACCTATCCTG 850

Qy    879 GAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTG 938
Db    851 GAATAAGACAGTCAGAAATGTGCTTGCAAGTGCATCGGCAGACAGCACTGTAGTTCTCTG 910

Qy    939 GGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAAC 998
Db    911 GGATCTGTCTGTGGGGAAGTCAGTAGCCAGGCTTACTGCACATACAGACAAGGTTCAGAC 970

Qy    999 ACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGT 1058
Db    971 TCTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGGTCCTATGACAAATCGGT 1030

Qy   1059 GGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCA 1118
Db   1031 GGCTCTGTATGACTGCAGAGATCCCAGCCAGAACCACCGCCAGTGGCGGTTCAGTGGGCA 1090

Qy   1119 GATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGA 1178
Db   1091 GATCGAGAGAGTGACCTGGAACCACTTCTCACCGTGTCATTTCTTGGCCAGTACAGATGA 1150

Qy   1179 CGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACA 1238
Db   1151 TGGCTTTGTCTATAACCTGGATGCACGTTCAGACAAGCCAATCTTTACACTTAATGCACA 1210

Qy   1239 CAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGC 1298
Db   1211 CAACGATGAAATCTCTGGTCTTGATCTTAGCAGTCAGATCAAAGGCTGCCTTGTGACTGC 1270

Qy   1299 TTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCA 1358
Db   1271 TTCCGCTGACAAATTTGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTCATTCA 1330

Qy   1359 TTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATT 1418
Db   1331 CTCTAGAGACATGAAAATGGGAGTTCTCTTTTGTTCCTCCTGCTGTCCTGACTTGCCATT 1390

Qy   1419 TATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTC 1478
Db   1391 TGTTTATGCCTTTGGAGGACAGAAAGAAGGGCTTCGGGTTTGGGACATAAGCACTGTGTC 1450

Qy   1479 TTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTC 1538
Db   1451 TTCAGTAAATGAAGCATTTGGAAGACGGGAGCGACTCGTTATAGGCGGCACGAAAGGTTT 1510

Qy   1539 ATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCA---GATACACCCATGGAGTCTTAATG 1595
Db   1511 ATCTGTTAGCGGCCCCTGTGGGAGCAGGAGTCCACAGCAGACACCAATGGAGTCCTGACA 1570

Qy   1596 AAGATCA-TCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAAT 1654
Db   1571 AAGATCAGTTTCCTGCTGTGCTTACCTTAACTCCG--TTTCAAAAGACTGGCCTAAAAAT 1628

Qy   1655 GTTC 1658
Db   1629 GTCC 1632
```

```
RESULT 4
BC003199
LOCUS       BC003199               1828 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 2310058A11 gene, clone MGC:6406
            IMAGE:3586496, mRNA, complete cds.
ACCESSION   BC003199
VERSION     BC003199.1  GI:13096811
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1828)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 10 Row: e Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
FEATURES             Location/Qualifiers
     source          1. .1828
                     /organism="Mus musculus"
```

```
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:6406 IMAGE:3586496"
                     /tissue_type="Mammary tumor metastatized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             82. .1587
                     /codon_start=1
                     /product="RIKEN cDNA 2310058A11 gene"
                     /protein_id="AAH03199.1"
                     /db_xref="GI:13096812"
                     /db_xref="LocusID:67485"
                     /translation="MNRSRQVTCVAWVRRGVAKETPDKVELSKEEVNRLIAEAKGKLQ
                     EEGGSEEEEAGNPSEDGMQSGPTQAPPRESLEDGDPQDDRTLDDDELAGYDLDNYDEE
                     DNPDAETIGESLLGLTVYGSNDQDPYVTLKDTEQYEHEDFLIKPTDNLIVCGRAEQEQ
                     CNLEVHVYNQEEESFYVHHDILLSAYPLSVEWLNFDPSPDASTGNYIAVGNMTPVIEV
                     WDLDIVDSLEPVFTLGSKLSKKKKKKGKKSSSAEGHTDAVLDLSWNKTVRNVLASASA
                     DSTVVLWDLSVGKSVARLTAHTDKVQTLQFHPFEAQTLISGSYDKSVALYDCRDPSQN
                     HRQWRFSGQIERVTWNHFSPCHFLASTDDGFVYNLDARSDKPIFTLNAHNDEISGLDL
                     SSQIKGCLVTASADKFVKIWDILGDRPSLIHSRDMKMGVLFCSSCCPDLPFVYAFGGQ
                     KEGLRVWDISTVSSVNEAFGRRERLVIGGTKGLSVSGPCGSRSPQQTPMES"
BASE COUNT      504 a    414 c    476 g    434 t
ORIGIN

  Query Match           62.2%;  Score 1152.4;  DB 10;  Length 1828;
  Best Local Similarity  82.7%;  Pred. No. 2e-302;
  Matches 1367;  Conservative    0;  Mismatches  276;  Indels    9;  Gaps    4;

Qy     11 CGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACG 70
Db      5 CGCGTGGGCGCGGAGCTCAAGAGTAGTGTCTTCCAGCTCCCAGGGCTCCAACCTGCGCCG 64

Qy     71 CCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCT 130
Db     65 GGTTCAAACCAAAGATCATGAACCGCAGCCGCCAAGTGACCTGCGTGGCCTGGGTCCGCC 124

Qy    131 GCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCC 190
Db    125 GCGGCGTGGCCAAAGAGACTCCAGACAAGGTAGAGCTAAGTAAAGAGGAAGTGAACCGTC 184

Qy    191 TCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGA 250
Db    185 TCATCGCTGAAGCAAAAGGAAAGTTGCAAGAAGAA---GGTGGGAGTGAAGAAGAGGAGG 241

Qy    251 CAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGC 310
Db    242 CAGGCAACCCTTCAGAAGATGGCATGCAGAGTGGCCCGACTCAGGCACCGCCTAGGGAGT 301

Qy    311 CCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGT 370
Db    302 CCCTGGAGGATGGGGATCCACAAGATGACCGCACACTGGACGATGACGAGCTGGCGGGGT 361

Qy    371 ACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTC 430
Db    362 ACGACTTAGATAATTATGACGAAGAGGATAACCCAGATGCTGAGACCATTGGCGAGTCTC 421

Qy    431 TCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATA 490
Db    422 TGTTGGGTCTTACGGTCTATGGCAGTAATGACCAGGATCCATATGTCACCTTGAAAGACA 481

Qy    491 CAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTG 550
Db    482 CGGAACAATATGAGCATGAAGATTTCTTGATTAAGCCGACTGACAACCTCATCGTTTGTG 541

Qy    551 GCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACT 610
Db    542 GCCGTGCTGAACAAGAGCAGTGCAATCTAGAGGTGCACGTTTACAATCAAGAAGAGGAGT 601

Qy    611 CTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGA 670
Db    602 CTTTCTACGTCCACCATGACATACTCCTGTCTGCATACCCTCTGAGTGTGGAATGGCTGA 661

Qy    671 ATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCC 730
Db    662 ATTTTGATCCTAGCCCAGATGCATCTACTGGGAATTACATTGCTGTGGGAAACATGACTC 721

Qy    731 CTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCG 790
Db    722 CTGTCATCGAGGTATGGGACCTGGATATAGTGGACTCTCTAGAACCAGTCTTCACACTGG 781

Qy    791 GAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGC 850
Db    782 GAAGCAAGCTTTCCAAAAAGAAGAAGAAGAAAGGGAAGAAGAGTTCCTCAGCAGAAGGAC 841

Qy    851 ATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTG 910
Db    842 ATACTGATGCTGTCCTTGACCTATCCTGGAATAAGACAGTCAGAAATGTGCTGGCAAGTG 901

Qy    911 CATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCC 970
Db    902 CATCGGCAGACAGCACTGTAGTTCTCTGGGATCTGTCTGTGGGGAAGTCAGTAGCCAGGC 961

Qy    971 TCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTC 1030
Db    962 TTACTGCACATACAGACAAGGTTCAGACTCTGCAGTTTCATCCATTTGAAGCACAGACTC 1021

Qy   1031 TGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAA 1090
Db   1022 TGATTTCTGGGTCCTATGACAAATCGGTGGCTCTGTATGACTGCAGAGATCCCAGCCAGA 1081

Qy   1091 GCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCAC 1150
Db   1082 ACCACCGCCAGTGGCGGTTCAGTGGGCAGATCGAGAGAGTGACCTGGAACCACTTCTCAC 1141

Qy   1151 CTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAG 1210
Db   1142 CGTGTCATTTCTTGGCCAGTACAGATGATGGCTTTGTGTATAACCTGGATGCACGTTCAG 1201

Qy   1211 ATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCA 1270
Db   1202 ACAAGCCAATCTTTACACTTAATGCACACAACGATGAAATCTCTGGTCTTGATCTTAGCA 1261

Qy   1271 GTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACA 1330
Db   1262 GTCAGATCAAAGGCTGCCTTGTGACTGCTTCCGCTGACAAATTTGTGAAGATCTGGGACA 1321

Qy   1331 TCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCT 1390
Db   1322 TCTTAGGAGATAGGCCAAGTCTCATTCACTCTAGAGACATGAAAATGGGAGTTCTCTTTT 1381

Qy   1391 GTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGC 1450
Db   1382 GTTCCTCCTGCTGTCCTGACTTGCCATTTGTTTATGCCTTTGGAGGACAGAAAGAAGGGC 1441

Qy   1451 TTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAGAGA 1510
Db   1442 TTCGGGTTTGGGACATAAGCACTGTGTCTTCAGTAAATGAAGCATTTGGAAGACGGGAGC 1501

Qy   1511 GGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGCAGCAGGAGCT 1570
Db   1502 GACTCGTTATAGGCGGCACGAAAGGTTTATCTGTTAGCGGCCCCTGTGGGAGCAGGAGTC 1561

Qy   1571 CA---GATACACCCATGGAGTCTTAATGAAGATCA-TCTAATTTCCTGCTTACCTTAACT 1626
Db   1562 CACAGCAGACACCAATGGAGTCCTGACAAAGATCAGTTTCCTGCTGTGCTTACCTTAACT 1621

Qy   1627 GGGAATTTTAAAAAGTTGGCCTAAAAATGTTC 1658
Db   1622 CCG--TTTCAAAAGACTGGCCTAAAAATGTCC 1651


RESULT 5
AC098211/c
```

```
LOCUS       AC098211              164664 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-208L16, *** SEQUENCING IN PROGRESS
            ***, 69 unordered pieces.
ACCESSION   AC098211
VERSION     AC098211.3  GI:21729861
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 164664)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164664)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 164664)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2002 this sequence version replaced gi:17973812.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIMR
                Center clone name: CH230-208L16
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 106583 bases at least Q40
                Consensus quality: 113107 bases at least Q30
                Consensus quality: 118472 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 69 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1313: contig of 1313 bp in length
            *     1314     1413: gap of unknown length
            *     1414     2511: contig of 1098 bp in length
            *     2512     2611: gap of unknown length
            *     2612     3846: contig of 1235 bp in length
            *     3847     3946: gap of unknown length
            *     3947     5519: contig of 1573 bp in length
            *     5520     5619: gap of unknown length
            *     5620     7133: contig of 1514 bp in length
            *     7134     7233: gap of unknown length
            *     7234     8594: contig of 1361 bp in length
            *     8595     8694: gap of unknown length
            *     8695     9703: contig of 1009 bp in length
            *     9704     9803: gap of unknown length
            *     9804    10945: contig of 1142 bp in length
            *    10946    11045: gap of unknown length
            *    11046    12242: contig of 1197 bp in length
            *    12243    12342: gap of unknown length
            *    12343    13881: contig of 1539 bp in length
            *    13882    13981: gap of unknown length
            *    13982    15230: contig of 1249 bp in length
            *    15231    15330: gap of unknown length
            *    15331    16364: contig of 1034 bp in length
            *    16365    16464: gap of unknown length
            *    16465    18075: contig of 1611 bp in length
            *    18076    18175: gap of unknown length
            *    18176    19710: contig of 1535 bp in length
            *    19711    19810: gap of unknown length
            *    19811    21291: contig of 1481 bp in length
            *    21292    21391: gap of unknown length
            *    21392    22816: contig of 1425 bp in length
            *    22817    22916: gap of unknown length
            *    22917    24259: contig of 1343 bp in length
            *    24260    24359: gap of unknown length
            *    24360    26369: contig of 2010 bp in length
            *    26370    26469: gap of unknown length
            *    26470    28190: contig of 1721 bp in length
            *    28191    28290: gap of unknown length
            *    28291    29566: contig of 1276 bp in length
            *    29567    29666: gap of unknown length
            *    29667    31464: contig of 1798 bp in length
            *    31465    31564: gap of unknown length
            *    31565    32733: contig of 1169 bp in length
            *    32734    32833: gap of unknown length
            *    32834    34340: contig of 1507 bp in length
            *    34341    34440: gap of unknown length
            *    34441    35861: contig of 1421 bp in length
            *    35862    35961: gap of unknown length
            *    35962    37213: contig of 1252 bp in length
            *    37214    37313: gap of unknown length
            *    37314    38899: contig of 1586 bp in length
            *    38900    38999: gap of unknown length
            *    39000    40786: contig of 1787 bp in length
```

```
*    40787    40886: gap of unknown length
*    40887    42840: contig of 1954 bp in length
*    42841    42940: gap of unknown length
*    42941    44820: contig of 1880 bp in length
*    44821    44920: gap of unknown length
*    44921    46296: contig of 1376 bp in length
*    46297    46396: gap of unknown length
*    46397    49196: contig of 2800 bp in length
*    49197    49296: gap of unknown length
*    49297    51055: contig of 1759 bp in length
*    51056    51155: gap of unknown length
*    51156    52468: contig of 1313 bp in length
*    52469    52568: gap of unknown length
*    52569    54246: contig of 1678 bp in length
*    54247    54346: gap of unknown length
*    54347    56244: contig of 1898 bp in length
*    56245    56344: gap of unknown length
*    56345    57759: contig of 1415 bp in length
*    57760    57859: gap of unknown length
*    57860    59092: contig of 1233 bp in length
*    59093    59192: gap of unknown length
*    59193    62231: contig of 3039 bp in length
*    62232    62331: gap of unknown length
*    62332    65033: contig of 2702 bp in length
*    65034    65133: gap of unknown length
*    65134    67535: contig of 2402 bp in length
*    67536    67635: gap of unknown length
*    67636    70030: contig of 2395 bp in length
*    70031    70130: gap of unknown length
*    70131    72994: contig of 2864 bp in length
*    72995    73094: gap of unknown length
*    73095    75682: contig of 2588 bp in length
*    75683    75782: gap of unknown length
*    75783    78532: contig of 2750 bp in length
*    78533    78632: gap of unknown length
*    78633    81157: contig of 2525 bp in length
*    81158    81257: gap of unknown length
*    81258    83896: contig of 2639 bp in length
*    83897    83996: gap of unknown length
*    83997    86151: contig of 2155 bp in length
*    86152    86251: gap of unknown length
*    86252    89033: contig of 2782 bp in length
*    89034    89133: gap of unknown length
*    89134    92970: contig of 3837 bp in length
*    92971    93070: gap of unknown length
*    93071    95027: contig of 1957 bp in length
*    95028    95127: gap of unknown length
*    95128    98206: contig of 3079 bp in length
*    98207    98306: gap of unknown length
*    98307   100157: contig of 1851 bp in length
*   100158   100257: gap of unknown length
*   100258   103213: contig of 2956 bp in length
*   103214   103313: gap of unknown length
*   103314   107610: contig of 4297 bp in length

  Query Match             53.2%;  Score 986.6;  DB 2;  Length 164664;
  Best Local Similarity   79.8%;  Pred. No. 5.3e-257;
  Matches 1257;  Conservative    0;  Mismatches  269;  Indels   49;  Gaps    6;

Qy      81 GAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGC 140
           ||  | ||||||  |||||||||||  |||||  |||||||||||  ||||||  ||||||||||
Db 157337 GAACATCATGAATCGCAGCCGCCAAGTGACCTGCGTGGCCTGCGTCCGCCGCGGCGTGGC 157278

Qy     141 CAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGA 200
           |||||||||  |||||||||||||||||  | ||||||  |||||  |||||  ||||||||||
Db 157277 CAAAGAGACGCCAGACAAGGTAGAGCTAAATAAAGAGGAAGTGAAACGTCTCATTGCTGA 157218

Qy     201 GGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCC 260
           |||||  ||  ||  ||  |||||  |||    |||||  |||||  |||||||||  ||||||||||
Db 157217 AGCAAAAGAAAAGTTACAAGAGGAA---GGTGGGAGTGAGGAAGAGGAGGCAGGCAGTCC 157161

Qy     261 TTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGA 320
           ||||||||||    ||||||  ||  ||  |||||||  |    |  ||  ||||  ||||||||
```

```
Db 157160 TTCAGAAGATG----GCAGAGAGCCCGGACCCAGGTCCAGTCTAGGGAGCAACTGGAGGA 157105

Qy    321 TGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGA 380
          |||  ||  |||  |  ||  |||    ||  ||  ||  ||  |||||  |||||  ||||||||
Db 157104 TGGGGATCCACAAGACGACCACACACTGGACGACGATGAACTGGCCAAGTACGAC----- 157050

Qy    381 TAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCT 440
                ||  ||  |    ||||||||||||||||  ||||||||||||  ||||||||
Db 157049 ----------GAAGAGGACAACCCAGATGCTGAGACCATTGGTGAATCTCTGTTGGGTCT 157000

Qy    441 TACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATA 500
          |||  |||||  ||  |||||  |||||  |  |||  ||||    ||  ||||||||  ||||||||
Db 156999 TACTGTCTAGGGCAGTAACGATCAGGGTCCATACGCCACCCTGAAAGACACAGAACAGTA 156940

Qy    501 TGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGA 560
          |||  ||||||||  |  ||||||||||  ||  |  ||  ||  ||  ||  ||||||||  |||||
Db 156939 TGAGCGTGAAGACTCCTTGATTAAACCGACTGGCAACCTCATCGTCTGTGGCCGTGCTGA 156880

Qy    561 ACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGT 620
          |||  |  |||||  ||  ||||||||||||  ||  ||  ||||||||||||  ||  |||||  |||||
Db 156879 ACAAGGGCAGTGTAACTTAGAGGTGCACGTGTACAATCAAGAAGAGGAGTCTTTCTATGT 156820

Qy    621 ACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCC 680
          ||||||||  ||||||  ||||||||||||  ||||||||||||||||||||||||||||||||||||
Db 156819 CCACCATGACATACTCCTGTCTGCATATC--CTGAGTGTGGAATGGCTGAATTTTGATCC 156762

Qy    681 TAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGA 740
          ||||||||||||  ||||||||||||||||||||||||||||||||||||||  |||  |||||
Db 156761 TAGCCCAGATGCGGCTACTGGAAATTACATTGCTGTGGGAAACATGACCCTTGTCATTGA 156702

Qy    741 AGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACT 800
          |  |||||||||  |||||  |||||||||  ||||  ||||||||||||||  |||  |  ||  ||
Db 156701 GGGGTGGGACCTGGATATCGTGGACTCTCTAGAACCAGTCTTCACACTGGGAGGCAAGCT 156642

Qy    801 TTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGC 860
          ||||||||||||||  ||||||||  ||  ||||||||  ||||||||||||||||||  |||||
Db 156641 TTCAAAAAAGAAGAAGAAGAAAGGGAAAAAGAGTTCGTCAGCAGAAGGGCATACTGATGC 156582

Qy    861 TGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGA 920
          ||||||||||||||||  ||||  ||||  |  |||||||  ||  |  |  ||||||||||||  ||
Db 156581 TGTCCTTGACCTTTCCTGGAGTAAGACAGTCAGAAACGTGCTGGTGAGTGCATCAGCAGA 156522

Qy    921 CAACACTGTAATTCTGT----GGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTG 976
          ||  |||||||||  ||||      ||||  ||  |  |||||||  |||||||  ||  |  |||
Db 156521 CAGCACTGTAATCCTGTGGGCGGGACCTGCCTGTGGGGAAGTCAGCAGCCAGGTTTACTG 156462

Qy    977 TACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTT 1036
          ||||||||||  ||||  |  ||  ||||||||||||||||||||||||||||||||||||||||||
Db 156461 CACACACAGACGAGGTTTAGACTCTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTC 156402

Qy   1037 CTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATC 1096
          ||||  |  |||||  ||  ||  ||||  ||||||  ||  |||  |  |  ||  ||  |
Db 156401 CTGGGTTCTATGACAAATCCGTGGCTCTGTACGACTGCAGAGATCCCAGCCAGAATCACC 156342

Qy   1097 GAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTC 1156
          |  |||||  |  ||||||||||||||||||||  ||||  |||||  |||||  |  |
Db 156341 GTCTGTGGTGGTTCAGTGGGCAGATAGAGAGGCTGACCTGGAACCACTTCTCACCGTCCC 156282

Qy   1157 ATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGC 1216
          |||||||  ||||||||||||||  ||||||||  ||||||  |||||||  ||  ||
Db 156281 ATTTCTTAGCCAGTACAGATGATGGCTTTGTGTATAATCTGGATGCGCGCTC-------- 156230

Qy   1217 CAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAA 1276
                    ||  ||||||||||  ||||||||||||||||||||||||||  ||||||  | |
Db 156229 ------------TACTGCACACAACGATGAAATCTCTGGTCTTGATCTGAGCAGTGAGA 156183

Qy   1277 TCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAG 1336
          ||||  |  |||  ||  ||||||||||||  |  |||||||  ||||  |||  ||  ||||||||||
Db 156182 TCAAAGACTGCCTTGTGACTGCTTCAACAGACAAATTTGTGACGATTTGAGACATCTTAG 156123

Qy   1337 GAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTT 1396
          |  |||||  ||||||||  ||||||||||  ||  ||||||||||||||||||  |||||||
Db 156122 GGGATAGACCAAGTCTCATTCATTCTAGAGATATGAAAATGGGAGTTCTCTTTTGTTCTT 156063
```

```
Qy   1397 CATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGG 1456
          | || || |||||||||||| ||| |||| |||||||||||| || ||||||||||||||||
Db 156062 CCTGCTGTCCTGATTTGCCCTTTGTTTACGCCTTTGGAGGACAGAAAGAAGGGCTTCGGG 156003

Qy   1457 TCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTG 1516
          | ||||| |||| ||| || |||||||||| ||||| |||||||||| ||||| || | |
Db 156002 TTTGGGACATAAACACTGTGTCTTCAGTAAACGAAGCGTTTGGAAGACGGGAGAGACTGG 155943

Qy   1517 TTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATA 1576
          || |||| ||  | | |  || |||| ||| |||||| | | | ||||||||| || | |
Db 155942 TTATTGGCAGCACGAAAGGTTTATCTGTTATTGGCCCCTGTAGGAGCAGGAGCCCACAGA 155883

Qy   1577 CACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTA 1636
          |||| |||||||| | |  |||||||  |  |    |||||||||||||| ||| ||| |
Db 155882 CACCAATGGAGTCCTGACAAAGATCAGTTCCTGGTGTGCTTACCTTAACTTGGAGTTTAA 155823

Qy   1637 AAAAGTTGGCCTAAA 1651
          |||| | | |  |||
Db 155822 AAAACTGGCCTAAAA 155808
```

RESULT 6
BC000067
LOCUS BC000067 807 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, nuclear phosphoprotein similar to S. cerevisiae PWP1, clone IMAGE:3505077, mRNA.
ACCESSION BC000067
VERSION BC000067.1 GI:13937531
KEYWORDS .
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 807)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 7 Row: f Column: 7
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5902033.
FEATURES Location/Qualifiers
source 1. .807
/organism="Homo sapiens"
/db_xref="LocusID:11137"
/db_xref="taxon:9606"
/clone="IMAGE:3505077"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
BASE COUNT 249 a 159 c 214 g 185 t
ORIGIN

Query Match 42.4%; Score 784.8; DB 9; Length 807;
Best Local Similarity 99.1%; Pred. No. 2.3e-202;

Matches 789; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy     37 CCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGC 96
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 CCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGC 69

Qy     97 AGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGAC 156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 AGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGAC 129

Qy    157 AAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTG 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 AAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTG 189

Qy    217 CAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATG 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 CAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATG 249

Qy    277 CAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGAT 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 CAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGAT 309

Qy    337 GACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAA 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 GACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAA 369

Qy    397 GGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGT 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 GGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGT 429

Qy    457 AATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTC 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 AATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTC 489

Qy    517 TTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAAT 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 TTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAAT 549

Qy    577 TTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTC 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 TTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTC 609

Qy    637 TTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCT 696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 TTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCT 669

Qy    697 ACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGAT 756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670 ACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGAT 729

Qy    757 ATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAA 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db    730 ATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAAAAAAAA 789

Qy    817 AAGAAAGGAAAGAAGA 832
          || |||  ||| || |
Db    790 AAAAAAAAAAAAAAAA 805
```

RESULT 7
AC007622/c
LOCUS AC007622 121700 bp DNA linear PRI 01-NOV-2000
DEFINITION Homo sapiens 12 BAC RP11-864J10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC007622
VERSION AC007622.28 GI:11067200
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 121700)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 121700)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 121700)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 1, 2000 this sequence version replaced gi:11038616.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

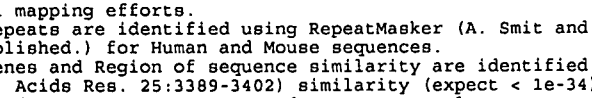
--- Summary Statistics ---

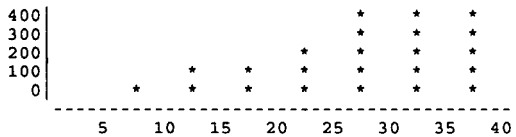
Contig length:	207416
Phrap values in estimate:	206947
Average error rate (BCM-Phrap estimate):	9.47342e-05
Fraction of Phrap values less than 40 :	0.0165888
Number of consensus changing edits:	31
Number of N's in consensus :	0

--- Consensus changing edits ---

Position	Original+Context	Edited+Context
22181	atggctcacg(n)ctgtaatccc	atggctcacg(c)ctgtaatccc
22487	ttttccatgt(n)ngtcttgaac	ttttccatgt(a)ggtcttgaac
22488	tttccatgtn(n)gtcttgaaca	tttccatgta(g)gtcttgaaca
22503	tgaacatatc(n)tattatgctt	tgaacatatc(t)tattatgctt
22523	tatttcatga(n)gttttatagg	tatttcatga(t)gttttatagg
22663	taagatctac(c)tanatnaaat	taagatctac(a)tagattaaat
22666	gatctaccta(n)atnaaatgca	gatctacata(g)attaaatgca
22669	ctacctanat(n)aaatgcacnn	ctacatagat(t)aaatgcacag
22678	tnaaatgcac(n)natgntaaat	ttaaatgcac(a)gatgttaaat
22679	naaatgcacn(n)atgntaaata	taaatgcaca(g)atgttaaata
22683	tgcacnnatg(n)taaatatacn	tgcacagatg(t)taaatataca
22693	ntaaatatac(n)acttaatgag	ttaaatatac(a)acttaatgag
34021	tgtctcaaaa(n)caaaaaaaaa	tgtctcaaaa(a)caaaaaaaaa
34471	tggcagtgca(n)gatggtttgc	tggcagtgca(a)gatggtttgc
44899	agtacagtgg(n)gtagctggct	agtacagtgg(t)gtagctggct
56139	acagtgcctg(n)cataattaga	acagtgcctg(g)cataattaga
73435	gatgtggtag(n)tgacttgttg	gatgtggtag(g)tgacttgttg
74766	tccacatttt(n)gggtttaaat	tccacatttt(g)gggtttaaat
94286	taacgatttt(n)aaagctgcag	taacgatttt(a)aaagctgcag
128065	gacagattgc(n)tgagctcggg	gacagattgc(t)tgagctcggg
169207	ctccatctta(n)nanaaanaaa	ctccatctta(a)aaaaaaaaaa
169208	tccatcttan(n)anaaanaaaa	tccatcttaa(a)aaaaaaaaaa
169210	catcttanna(n)aaanaaaaan	catcttaaaa(a)aaaaaaaaaa
169214	ttannanaaa(n)aaaaanaaga	ttaaaaaaaa(a)aaaaaaaaga
169220	naaanaaaaa(n)aagacaaata	aaaaaaaaaa(a)aagacaaata
169276	aaatgaatat(n)ttgttctgct	aaatgaatat(a)ttgttctgct
191868	cccaacacac(n)anagttaata	ccaacacac(t)atagttaata
191870	caacacacna(n)agttaatatc	caacacacta(t)agttaatatc
198887	atttctagtg(n)tccaatcccc	atttctagtg(t)tccaatcccc
203149	aatggtttgt(n)cacattattt	aatggtttgt(a)cacattattt
207416	ctcatggaat(n).	ctcatggaat(c). . . . .

--- Distribution of Quality < 40 Bases ---






```
          Version: 1.01  qxfo.
FEATURES             Location/Qualifiers
     source          1. .121700
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-864J10"
     repeat_region   complement(1087. .1398)
                     /rpt_family="AluSx"
     repeat_region   3261. .3665
                     /rpt_family="MER61C"
     repeat_region   complement(4390. .4440)
                     /rpt_family="MIR"
     STS             4662. .4839
                     /standard_name="11.126A/11.126B"
                     /db_xref="GDB:385913"
     repeat_region   4692. .4742
                     /rpt_family="(TG)n"
     repeat_region   4764. .4895
                     /rpt_family="L1MC4"
     repeat_region   4901. .5177
                     /rpt_family="AluJo"
     repeat_region   5178. .5214
                     /rpt_family="(CAAT)n"
     repeat_region   5215. .5453
                     /rpt_family="AluJo"
     repeat_region   5483. .6031
                     /rpt_family="L1MC4"
     repeat_region   6032. .6338
                     /rpt_family="AluSx"
     repeat_region   6349. .6445

  Query Match             20.2%;  Score 373.4;  DB 9;  Length 121700;
  Best Local Similarity   99.7%;  Pred. No. 5.2e-90;
  Matches  374;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1479 TTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTC 1538
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80178 TTTAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTC 80119

Qy   1539 ATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAG 1598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80118 ATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAG 80059

Qy   1599 ATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTC 1658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  80058 ATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTC 79999

Qy   1659 CATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCC 1718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79998 CATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCC 79939

Qy   1719 TTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGAT 1778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79938 TTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGAT 79879

Qy   1779 GGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAG 1838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  79878 GGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAG 79819

Qy   1839 TAATTCCTTAAACAT 1853
           ||||||||||||||
```

```
Db  79818 TAATTCCTTAAACAT 79804


RESULT 8
AC010187/c
LOCUS       AC010187               334597 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-389O9, WORKING DRAFT
            SEQUENCE, 40 unordered pieces.
ACCESSION   AC010187
VERSION     AC010187.14  GI:20279317
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334597)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 334597)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 334597)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 24, 2002 this sequence version replaced gi:15667986.
            -------------- Genome Center
                Center: Baylor College of Medicine
```

```
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: HMKV
    Center clone name: RP11-389O9
-------------- Summary Statistics
    Sequencing vector: M13;
    Chemistry: Dye-primer Bodipy: 77% of reads
    Chemistry: Dye-terminator Big Dye: 23% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 299089 bases at least Q40
    Consensus quality: 323681 bases at least Q30
    Consensus quality: 338387 bases at least Q20
    Estimated insert size: 349203; sum-of-contigs estimation
    Estimated insert size: 503842; agarose-fp estimation
    Quality coverage: 2x in Q20 bases; agarose-fp estimation
    Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     2157: contig of 2157 bp in length
*    2158     2257: gap of unknown length
*    2258     5490: contig of 3233 bp in length
*    5491     5590: gap of unknown length
*    5591     8466: contig of 2876 bp in length
*    8467     8566: gap of unknown length
*    8567    10661: contig of 2095 bp in length
*   10662    10761: gap of unknown length
*   10762    13041: contig of 2280 bp in length
*   13042    13141: gap of unknown length
*   13142    16011: contig of 2870 bp in length
*   16012    16111: gap of unknown length
*   16112    18931: contig of 2820 bp in length
*   18932    19031: gap of unknown length
*   19032    22675: contig of 3644 bp in length
*   22676    22775: gap of unknown length
*   22776    25400: contig of 2625 bp in length
*   25401    25500: gap of unknown length
*   25501    29388: contig of 3888 bp in length
*   29389    29488: gap of unknown length
*   29489    32970: contig of 3482 bp in length
*   32971    33070: gap of unknown length
*   33071    36207: contig of 3137 bp in length
*   36208    36307: gap of unknown length
*   36308    39878: contig of 3571 bp in length
*   39879    39978: gap of unknown length
*   39979    43312: contig of 3334 bp in length
*   43313    43412: gap of unknown length
*   43413    47414: contig of 4002 bp in length
*   47415    47514: gap of unknown length
*   47515    50164: contig of 2650 bp in length
*   50165    50264: gap of unknown length
*   50265    54914: contig of 4650 bp in length
*   54915    55014: gap of unknown length
*   55015    57456: contig of 2442 bp in length
*   57457    57556: gap of unknown length
*   57557    60864: contig of 3308 bp in length
*   60865    60964: gap of unknown length
*   60965    65163: contig of 4199 bp in length
*   65164    65263: gap of unknown length
*   65264    70398: contig of 5135 bp in length
*   70399    70498: gap of unknown length
*   70499    74623: contig of 4125 bp in length
*   74624    74723: gap of unknown length
*   74724    78535: contig of 3812 bp in length
*   78536    78635: gap of unknown length
*   78636    82429: contig of 3794 bp in length
*   82430    82529: gap of unknown length
*   82530    88169: contig of 5640 bp in length
*   88170    88269: gap of unknown length
*   88270    94292: contig of 6023 bp in length
*   94293    94392: gap of unknown length
*   94393   101749: contig of 7357 bp in length
*  101750   101849: gap of unknown length
*  101850   109260: contig of 7411 bp in length
*  109261   109360: gap of unknown length
*  109361   116775: contig of 7415 bp in length
*  116776   116875: gap of unknown length
*  116876   123551: contig of 6676 bp in length
*  123552   123651: gap of unknown length
*  123652   127603: contig of 3952 bp in length
*  127604   127703: gap of unknown length
*  127704   136510: contig of 8807 bp in length
*  136511   136610: gap of unknown length
*  136611   144218: contig of 7608 bp in length
*  144219   144318: gap of unknown length
*  144319   158470: contig of 14152 bp in length
*  158471   158570: gap of unknown length
*  158571   171113: contig of 12543 bp in length
*  171114   171213: gap of unknown length
*  171214   188563: contig of 17350 bp in length
*  188564   188663: gap of unknown length
*  188664   222888: contig of 34225 bp in length
*  222889   222988: gap of unknown length
*  222989   257090: contig of 34102 bp in length
*  257091   257190: gap of unknown length
*  257191   295336: contig of 38146 bp in length
*  295337   295436: gap of unknown length
*  295437   334597: contig of 39161 bp in length.
FEATURES             Location/Qualifiers
     source          1. .334597
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-389O9"
BASE COUNT    90733 a  73766 c  75238 g  90890 t   3970 others
ORIGIN

  Query Match            20.2%;  Score 373.4;  DB 2;  Length 334597;
  Best Local Similarity   99.7%;  Pred. No. 6e-90;
  Matches  374;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1479 TTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTC 1538
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48142 TTTAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTC 48083

Qy   1539 ATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAG 1598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48082 ATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAG 48023

Qy   1599 ATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTC 1658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  48022 ATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTC 47963

Qy   1659 CATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCC 1718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47962 CATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCC 47903

Qy   1719 TTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGAT 1778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47902 TTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGAT 47843

Qy   1779 GGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAG 1838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  47842 GGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAG 47783
```

```
Qy   1839 TAATTCCTTAAACAT 1853
          ||||||||||||||
Db  47782 TAATTCCTTAAACAT 47768


RESULT 9
HSXS153
LOCUS       HSXS153                  400 bp    mRNA    linear   PRI 10-APR-1997
DEFINITION  H.sapiens (xs153) mRNA, 400bp.
ACCESSION   Z36805
VERSION     Z36805.1  GI:533924
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Mueller-Pillasch,F., Gress,T., Lehrach,H. and Adler,G.
  TITLE     Differential gene expression in pancreatic cancer. Use of an
            automated approach for the large scale isolation and
            characterisation of cDNA clones containing differentially expressed
            sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 400)
  AUTHORS   Gress,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1994) Gress T., University of Ulm, Department of
            Internal Medicine I, Robert Koch Str.8, 89081 Ulm, Germany, 89081
REFERENCE   3  (bases 1 to 400)
  AUTHORS   Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F.,
            Zehetner,G., Friess,H., Buchler,M., Adler,G. and Lehrach,H.
  TITLE     A pancreatic cancer-specific expression profile
  JOURNAL   Oncogene 13 (8), 1819-1830 (1996)
  MEDLINE   97050791
   PUBMED   8895530
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="xs153"
                     /cell_line="PATU 8988s"
                     /cell_type="ductal adenocarcinoma"
                     /tissue_type="pancreas"
                     /clone_lib="xs cDNA lib, pancreatic cancer cell line PATU
                     8988s"
                     /dev_stage="adult"
     misc_feature    1. .382
                     /product="56 kDa protein"
                     /note="95% homology to human 56 kDa protein, GenEMBL-locus
                     gb_pr:hum56kdapr residues 419-797"
BASE COUNT      113 a     78 c     90 g    115 t      4 others
ORIGIN

  Query Match             18.2%;  Score 337.6;  DB 9;  Length 400;
  Best Local Similarity   94.8%;  Pred. No. 1.3e-80;
  Matches  379;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    3;

Qy    419 TTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTA 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTA 60

Qy    479 CTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATC 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     61 CTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATACTC 120

Qy    539 TTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATC 598
          ||||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTATAGGGTGTGGGCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATC 180

Qy    599 AAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTG 658
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    181 AAGAAGAAGACTCTTTTTATGTACACCAAGATATACTCTTGTCTGCATATCCTCTGAGTG 240

Qy    659 TGGAATGGCTGAATT-TTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTA 717
          ||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGAATGGCTGAATTGTNGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTA 300

Qy    718 GGAAACAT-GACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCC 776
          ||||||||  ||||||||||||||||||||  ||||||||||||||||||||||||| |
Db    301 GGAAACATCCCACCCCTCTTATTGAAGTGTGGGACCCCGATATAGTGGACTCTTTAGAGNC 360

Qy    777 AGTCTTCACACTC-GGAAGTAAACTTTCAAAAAAGAAGAA 815
          |||||||||||||  ||||||  |||  ||||  || ||||
Db    361 AGTCTTCACACTCGGGAAGTTNACTNCCAAACCAGCAGAA 400


RESULT 10
AC109397
LOCUS       AC109397              112300 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-261M6, *** SEQUENCING IN PROGRESS
            ***, 52 unordered pieces.
ACCESSION   AC109397
VERSION     AC109397.3  GI:21465200
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 112300)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112300)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Human Genome Sequencing Center, Department
```

```
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 112300)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jun 19, 2002 this sequence version replaced gi:18846849.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GPZA
              Center clone name: CH230-261M6
          -------------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 67877 bases at least Q40
              Consensus quality: 71488 bases at least Q30
              Consensus quality: 74468 bases at least Q20
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 52 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1392: contig of 1392 bp in length
          *     1393     1492: gap of unknown length
          *     1493     2616: contig of 1124 bp in length
          *     2617     2716: gap of unknown length
          *     2717     4137: contig of 1421 bp in length
          *     4138     4237: gap of unknown length
          *     4238     5861: contig of 1624 bp in length
          *     5862     5961: gap of unknown length
          *     5962     7254: contig of 1293 bp in length
          *     7255     7354: gap of unknown length
          *     7355     8434: contig of 1080 bp in length
          *     8435     8534: gap of unknown length
          *     8535     9591: contig of 1057 bp in length
          *     9592     9691: gap of unknown length
          *     9692    10738: contig of 1047 bp in length
          *    10739    10838: gap of unknown length
          *    10839    11893: contig of 1055 bp in length
          *    11894    11993: gap of unknown length
          *    11994    13704: contig of 1711 bp in length
          *    13705    13804: gap of unknown length
          *    13805    15599: contig of 1795 bp in length
          *    15600    15699: gap of unknown length
          *    15700    16725: contig of 1026 bp in length
          *    16726    16825: gap of unknown length
          *    16826    18609: contig of 1784 bp in length
          *    18610    18709: gap of unknown length
          *    18710    19938: contig of 1229 bp in length
          *    19939    20038: gap of unknown length
          *    20039    21273: contig of 1235 bp in length
          *    21274    21373: gap of unknown length
          *    21374    23028: contig of 1655 bp in length
          *    23029    23128: gap of unknown length
          *    23129    24597: contig of 1469 bp in length
          *    24598    24697: gap of unknown length
          *    24698    27845: contig of 3148 bp in length
          *    27846    27945: gap of unknown length
          *    27946    29017: contig of 1072 bp in length
          *    29018    29117: gap of unknown length
          *    29118    31039: contig of 1922 bp in length
          *    31040    31139: gap of unknown length
          *    31140    32589: contig of 1450 bp in length
          *    32590    32689: gap of unknown length
          *    32690    35113: contig of 2424 bp in length
          *    35114    35213: gap of unknown length
          *    35214    37124: contig of 1911 bp in length
          *    37125    37224: gap of unknown length
          *    37225    39052: contig of 1828 bp in length
          *    39053    39152: gap of unknown length
          *    39153    40880: contig of 1728 bp in length
          *    40881    40980: gap of unknown length
          *    40981    42452: contig of 1472 bp in length
          *    42453    42552: gap of unknown length
          *    42553    44335: contig of 1783 bp in length
          *    44336    44435: gap of unknown length
          *    44436    45624: contig of 1189 bp in length
          *    45625    45724: gap of unknown length
          *    45725    48195: contig of 2471 bp in length
          *    48196    48295: gap of unknown length
          *    48296    49958: contig of 1663 bp in length
          *    49959    50058: gap of unknown length
          *    50059    52259: contig of 2201 bp in length
          *    52260    52359: gap of unknown length
          *    52360    53726: contig of 1367 bp in length
          *    53727    53826: gap of unknown length
          *    53827    55782: contig of 1956 bp in length
          *    55783    55882: gap of unknown length
          *    55883    57544: contig of 1662 bp in length
          *    57545    57644: gap of unknown length
          *    57645    59237: contig of 1593 bp in length
          *    59238    59337: gap of unknown length
          *    59338    61091: contig of 1754 bp in length
          *    61092    61191: gap of unknown length
          *    61192    62543: contig of 1352 bp in length
          *    62544    62643: gap of unknown length
          *    62644    65077: contig of 2434 bp in length
          *    65078    65177: gap of unknown length
          *    65178    67369: contig of 2192 bp in length
          *    67370    67469: gap of unknown length
          *    67470    69466: contig of 1997 bp in length
          *    69467    69566: gap of unknown length
          *    69567    72140: contig of 2574 bp in length
          *    72141    72240: gap of unknown length
          *    72241    74313: contig of 2073 bp in length
          *    74314    74413: gap of unknown length
          *    74414    77865: contig of 3452 bp in length
          *    77866    77965: gap of unknown length
          *    77966    81535: contig of 3570 bp in length
          *    81536    81635: gap of unknown length
          *    81636    85369: contig of 3734 bp in length
          *    85370    85469: gap of unknown length
          *    85470    88583: contig of 3114 bp in length
          *    88584    88683: gap of unknown length
          *    88684    92585: contig of 3902 bp in length
          *    92586    92685: gap of unknown length
          *    92686    96502: contig of 3817 bp in length
          *    96503    96602: gap of unknown length
          *    96603   100074: contig of 3472 bp in length
          *   100075   100174: gap of unknown length
          *   100175   102758: contig of 2584 bp in length
          *   102759   102858: gap of unknown length
          *   102859   108435: contig of 5577 bp in length
          *   108436   108535: gap of unknown length
          *   108536   112300: contig of 3765 bp in length.
FEATURES             Location/Qualifiers
     source          1. .112300
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"

  Query Match             15.4%;  Score 285;  DB 2;  Length 112300;
  Best Local Similarity   77.6%;  Pred. No. 6.1e-66;
  Matches  384;  Conservative    0;  Mismatches  105;  Indels    6;  Gaps    3;
```

```
Qy   1067 ATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGA 1126
          ||||||||  |    |       |||    | |   | ||||  |||||||| |||||  |
Db  40374 ATGACTGCAGTGAACACATCCAAAATTACCTCCTATGGCACTTCAGTGGCCAGATCAAAT 40433

Qy   1127 GAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTG 1186
           ||||||   ||| || || ||||| |||||| ||||||||||||||||||| || ||||
Db  40434 AAGTGACCGAGAACCATTTCTCACCATGTCATATCTTGGCCAGTACAGATGATGGTTTTG 40493

Qy   1187 TATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATG 1246
          ||||||||  |||||  |||||||||| |||   |||||||||||||||||||||  |  |||
Db  40494 TATATAATCTGGATACACGTTCAGAGAAGTTGATTTTTACACTTAATGCACACGACTATG 40553

Qy   1247 AAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTG 1306
          ||||  |||||||||||||| ||||||  | |||||  |||||  ||  |||||||   ||  | |
Db  40554 AAATGTCTGGTCTTGATCTGAGCAGTTAGATCAAAGGCTGCCTTGTGACTG--TCCACAG 40611

Qy   1307 ACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGG 1366
          ||||  |  |||||||||||||| ||||||||||| ||||||||||   |||| |||||  |
Db  40612 ACAATTTTGTGAAGATCTGGGATATCTTAGGAGAAAGGCCAAGTCTCATTCACTCTAGAG 40671

Qy   1367 ACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATG 1426
          ||||  ||||||  ||||||||||   ||| ||| | |  |||||||| |||||  ||||||
Db  40672 ACATTAAAATGAGAGTTCTCTT---TTCATCAAGCTTTCCTGATTTGTCATTTGTTTATG 40728

Qy   1427 CCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAA 1486
          ||||||||||  ||  |||||||||||| ||||  ||||||  |||||||  ||||||||||||
Db  40729 CCTTTGGAGGACAGAAAGAAGGGCTTTGGGTTTGGGAT-TAAGCACTGTCTCTTCAGTAA 40787

Qy   1487 ATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTA 1546
          ||||||  ||||||||||  |||||  ||  ||||  |      |  |  ||     ||||
Db  40788 ATGAAGTGTTTGGAAGACATGAGAGACTGATTATTGGCAACAGCATATTTTGGGTAATTA 40847

Qy   1547 GTGGCCCTTTTGGCA 1561
          ||||||  |  ||  ||
Db  40848 GTGGCCTCTATGCCA 40862
```

```
RESULT 11
G06393
LOCUS       G06393                   260 bp    DNA     linear   STS 19-OCT-1995
DEFINITION  human STS WI-7102, sequence tagged site.
ACCESSION   G06393
VERSION     G06393.1  GI:859638
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: ATTTCCTGCTTACCTTAACTGGG
            Primer B: CTTTTTTAAGAAATGCAACAAGAGC
            STS size: 207
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Prepared with primer pairs derived from L07758 -- Unigene.
FEATURES              Location/Qualifiers
     source           1. .260
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="729_D_6; 744_G_5; 909_G_6; 925_G_1; 948_F_3;
                      959_A_1; 720_G_10; 786_F_12; 969_C_12"
     STS              14. .220
     primer_bind      14. .36
     primer_bind      complement(196. .220)
BASE COUNT       79 a     49 c     48 g     84 t
ORIGIN
```

```
  Query Match             14.0%;  Score 260;  DB 11;  Length 260;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-59;
  Matches  260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1594 TGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAA 1653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAA 60

Qy   1654 TGTTCCATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGAC 1713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTTCCATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGAC 120

Qy   1714 ATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGA 1773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGA 180

Qy   1774 TGGATGGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTA 1833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGATGGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTA 240

Qy   1834 AAAAGTAATTCCTTAAACAT 1853
          ||||||||||||||||||||
Db    241 AAAAGTAATTCCTTAAACAT 260
```

```
RESULT 12
HSXS152
LOCUS       HSXS152                  250 bp    mRNA    linear   PRI 10-APR-1997
DEFINITION  H.sapiens (xs152) mRNA, 250bp.
ACCESSION   Z36804
VERSION     Z36804.1  GI:533911
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Mueller-Pillasch,F., Gress,T., Lehrach,H. and Adler,G.
  TITLE     Differential gene expression in pancreatic cancer. Use of an
            automated approach for the large scale isolation and
            characterisation of cDNA clones containing differentially expressed
            sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 250)
  AUTHORS   Gress,T.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (16-AUG-1994) Gress T., University of Ulm, Department of
            Internal Medicine I, Robert Koch Str.8, 89081 Ulm, Germany, 89081
REFERENCE   3  (bases 1 to 250)
  AUTHORS   Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F.,
            Zehetner,G., Friess,H., Buchler,M., Adler,G. and Lehrach,H.
  TITLE     A pancreatic cancer-specific expression profile
  JOURNAL   Oncogene 13 (8), 1819-1830 (1996)
  MEDLINE   97050791
   PUBMED   8895530
FEATURES             Location/Qualifiers
     source          1. .250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="xs152"
                     /cell_line="PATU 8988s"
                     /cell_type="ductal adenocarcinoma"
                     /tissue_type="pancreas"
                     /clone_lib="xs cDNA lib, pancreatic cancer cell line PATU
                     8988s"
                     /dev_stage="adult"
     misc_feature    1. .250
                     /product="56 kDa protein"
                     /note="98% homology to human 56 kDa protein, GenEMBL-locus
                     gb_pr:hum56kdapr residues 419-669"
BASE COUNT       70 a     45 c     56 g     75 t      4 others
ORIGIN

  Query Match             13.1%;  Score 242.8;  DB 9;  Length 250;
  Best Local Similarity   97.6%;  Pred. No. 8e-55;
  Matches  244;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     418 CTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTT 477
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTT 60

Qy     478 ACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAAT 537
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      61 ACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGNGATAAT 120

Qy     538 CTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAAT 597
           |||||||| || ||||||||||||||||||| ||||||||||||||||||||||||| |
Db     121 CTTATAGTGTGNGGCCGAGCTGAACAGGACCAGCNCAATTTAGAGGTGCATGTTTATANT 180

Qy     598 CAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGT 657
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGT 240

Qy     658 GTGGAATGGC 667
           ||||||||||
Db     241 GTGGAATGGC 250


RESULT 13
AC010187
LOCUS       AC010187               334597 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-389O9, WORKING DRAFT
            SEQUENCE, 40 unordered pieces.
ACCESSION   AC010187
VERSION     AC010187.14  GI:20279317
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334597)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 334597)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 334597)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 24, 2002 this sequence version replaced gi:15667986.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMKV
                Center clone name: RP11-389O9
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 77% of reads
                Chemistry: Dye-terminator Big Dye: 23% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 299089 bases at least Q40
                Consensus quality: 323681 bases at least Q30
                Consensus quality: 338387 bases at least Q20
                Estimated insert size: 349203; sum-of-contigs estimation
                Estimated insert size: 503842; agarose-fp estimation
                Quality coverage: 2x in Q20 bases; agarose-fp estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 40 contigs. The true order of the pieces
```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     2157: contig of 2157 bp in length
*    2158     2257: gap of unknown length
*    2258     5490: contig of 3233 bp in length
*    5491     5590: gap of unknown length
*    5591     8466: contig of 2876 bp in length
*    8467     8566: gap of unknown length
*    8567    10661: contig of 2095 bp in length
*   10662    10761: gap of unknown length
*   10762    13041: contig of 2280 bp in length
*   13042    13141: gap of unknown length
*   13142    16011: contig of 2870 bp in length
*   16012    16111: gap of unknown length
*   16112    18931: contig of 2820 bp in length
*   18932    19031: gap of unknown length
*   19032    22675: contig of 3644 bp in length
*   22676    22775: gap of unknown length
*   22776    25400: contig of 2625 bp in length
*   25401    25500: gap of unknown length
*   25501    29388: contig of 3888 bp in length
*   29389    29488: gap of unknown length
*   29489    32970: contig of 3482 bp in length
*   32971    33070: gap of unknown length
*   33071    36207: contig of 3137 bp in length
*   36208    36307: gap of unknown length
*   36308    39878: contig of 3571 bp in length
*   39879    39978: gap of unknown length
*   39979    43312: contig of 3334 bp in length
*   43313    43412: gap of unknown length
*   43413    47414: contig of 4002 bp in length
*   47415    47514: gap of unknown length
*   47515    50164: contig of 2650 bp in length
*   50165    50264: gap of unknown length
*   50265    54914: contig of 4650 bp in length
*   54915    55014: gap of unknown length
*   55015    57456: contig of 2442 bp in length
*   57457    57556: gap of unknown length
*   57557    60864: contig of 3308 bp in length
*   60865    60964: gap of unknown length
*   60965    65163: contig of 4199 bp in length
*   65164    65263: gap of unknown length
*   65264    70398: contig of 5135 bp in length
*   70399    70498: gap of unknown length
*   70499    74623: contig of 4125 bp in length
*   74624    74723: gap of unknown length
*   74724    78535: contig of 3812 bp in length
*   78536    78635: gap of unknown length
*   78636    82429: contig of 3794 bp in length
*   82430    82529: gap of unknown length
*   82530    88169: contig of 5640 bp in length
*   88170    88269: gap of unknown length
*   88270    94292: contig of 6023 bp in length
*   94293    94392: gap of unknown length
*   94393   101749: contig of 7357 bp in length
*  101750   101849: gap of unknown length
*  101850   109260: contig of 7411 bp in length
*  109261   109360: gap of unknown length
*  109361   116775: contig of 7415 bp in length
*  116776   116875: gap of unknown length
*  116876   123551: contig of 6676 bp in length
*  123552   123651: gap of unknown length
*  123652   127603: contig of 3952 bp in length
*  127604   127703: gap of unknown length
*  127704   136510: contig of 8807 bp in length
*  136511   136610: gap of unknown length
*  136611   144218: contig of 7608 bp in length
*  144219   144318: gap of unknown length
*  144319   158470: contig of 14152 bp in length
*  158471   158570: gap of unknown length
*  158571   171113: contig of 12543 bp in length
*  171114   171213: gap of unknown length
*  171214   188563: contig of 17350 bp in length
*  188564   188663: gap of unknown length
*  188664   222888: contig of 34225 bp in length
*  222889   222988: gap of unknown length
*  222989   257090: contig of 34102 bp in length
*  257091   257190: gap of unknown length
*  257191   295336: contig of 38146 bp in length
*  295337   295436: gap of unknown length
*  295437   334597: contig of 39161 bp in length.
FEATURES             Location/Qualifiers
     source          1. .334597
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-389O9"
BASE COUNT    90733 a  73766 c  75238 g  90890 t   3970 others
ORIGIN

  Query Match            10.2%;  Score 188.4;  DB 2;  Length 334597;
  Best Local Similarity   99.5%;  Pred. No. 1.4e-39;
  Matches  189;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    219 AGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCA 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  76886 AGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCA 76945

Qy    279 GAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGA 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  76946 GAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGA 77005

Qy    339 CAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGG 398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  77006 CAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGG 77065

Qy    399 TGACCCAGAT 408
          |||||||| |
Db  77066 TGACCCAGGT 77075


RESULT 14
AX410079/c
LOCUS       AX410079                150 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2726 from Patent WO0229103.
ACCESSION   AX410079
VERSION     AX410079.1  GI:21442784
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2726 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. N92948"
BASE COUNT       44 a     33 c     28 g     45 t
ORIGIN

  Query Match             8.1%;  Score 150;  DB 6;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-29;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    660 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 91
```

```
Qy     720 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 31

Qy     780 CTTCACACTCGGAAGTAAACTTTCAAAAAA 809
           |||||||||||||||||||||||||||||
Db      30 CTTCACACTCGGAAGTAAACTTTCAAAAAA 1


RESULT 15
SPBC1711
LOCUS       SPBC1711               38000 bp    DNA     linear   PLN 19-APR-2000
DEFINITION  S.pombe chromosome II cosmid c1711.
ACCESSION   AL353012
VERSION     AL353012.1  GI:7630158
KEYWORDS    60s ribosomal protein l2; confirmed intron; conserved hypothetical
            protein; conserved hypothetical transmembrane protein; dipeptidyl
            peptidase; his2; histidinol dehydrogenase; mating-type m-specific
            polypeptide mc; mating-type m-specific polypeptide mi; matmc;
            matmi; meiotic recombination protein; methylenetetrahydrofolate
            dehydrogenase [nad+]; nexin sorting protein; nuclear pore
            complex-associated; nuclear pore protein; periodic tryptophan
            protein 1 homolog; putative pre-mrna splicing factor atp-dependent
            rna helicase; PX domain; rec15; rpl4; serine-rich repeat protein;
            signal peptide; similar to yeast YKR088C; Tf2-type LTR; WD repeat
            protein.
SOURCE      Schizosaccharomyces pombe.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 38000)
  AUTHORS   Aert,R., Volckaert,G., Wood,V., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2000) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and Laboratory of Gene Technology, Katholieke Universiteit Leuven,
            Kardinaal Mercierlaan 92, B-3001 Leuven, Belgium
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre.  The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
            pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c1711 is overlapped at the 3' end by cosmid c17G9, EMBL
            entry SPBC17G9, accession number AL109850.
FEATURES             Location/Qualifiers
     source          1. .38000
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="II"
                     /map="IIR"
                     /clone="cosmid c1711"
     misc_feature    1472. .2794
                     /note="similarity to region of SPBC23G7 S. pombe
                     chromosome 2"
     gene            complement(1591. .1719)
                     /gene="matmi"
                     /note="SPBC1711.01c"
     CDS             complement(1591. .1719)
                     /gene="matmi"
                     /note="SPBC1711.01c, len:43"
                     /codon_start=1
                     /label=matmi
                     /product="mating-type m-specific polypeptide mi."
                     /protein_id="CAB88231.1"
                     /db_xref="GI:7630159"
                     /translation="MSAEDLFTIQILCDQIELKLASIVINSNIKLQLKRKKKTQQL"
     gene            1854. .2399
                     /gene="matmc"
                     /note="SPBC1711.02"
     CDS             1854. .2399
                     /gene="matmc"
                     /note="SPBC1711.02, len:182"
                     /codon_start=1
                     /label=matmc
                     /product="mating-type m-specific polypeptide mc."
                     /protein_id="CAB88232.1"
                     /db_xref="GI:7630160"
                     /translation="MDSHQELSAGSPISYDFLDPDWCFKRYLTKDALHSIETGKGAAY
                     FVPDGFTPILIPNSQSYLLDGNSAQLPRPQPISFTLDQCKVPGYILKSLRKDTTSTER
                     TPRPPNAFILYRKEKHATLLKSNPSINNSQVSKLVGEMWRNESKEVRMRYFKMSEFYK
                     AQHQKMYPGYKYQPRKNKVKR"
     misc_feature    2160. .2366
                     /gene="matmc"
                     /note="Match to PF00505 HMG_box, HMG (high mobility group)
                     box Score 118.06"
     misc_feature    complement(5424. .6126)
                     /note="region of similarity to SPBC23G7 S. pombe
                     chromosome 2"
     gene            join(7368. .7392,7432. .7558,7607. .7881,8087. .8248,
                     8294. .8481)
                     /gene="SPBC1711.03"
     CDS             join(7368. .7392,7432. .7558,7607. .7881,8087. .8248,
                     8294. .8481)
                     /gene="SPBC1711.03"
                     /note="SPBC1711.03, len:259, SIMILARITY:Arabidopsis
                     thaliana, CAB78302, hypothetical 27.9 kd protein., (246
                     aa), fasta scores: opt: 568, E():4.3e-31, (38.5% identity
                     in 252 aa)"
                     /codon_start=1
                     /label=SPBC1711.03
                     /product="conserved hypothetical transmembrane protein"
                     /protein_id="CAB88233.1"
                     /db_xref="GI:7630161"
                     /translation="MQKLLLDPALRNWVLLPIMFVMILIGILRHNATILLQSSPKKLS
                     KEEIREQRLLQRAYALRACSNSLLPESIEARKCFLIESLKSGKYLKPVDPNAPKAANP
                     LMDDKTLEGLMESMKGNMLMVVPQTIIMTWINEFFSGFILLKLPFPLTLRFKSIFQSG
                     VATQDLDVQWVSSISWYFLNLFGLKSVYALLLGENNAASNATNEMGMAGFSSAAATAQ
                     LIQPGQDISKMMLSEAENVQILKNESLLVDVEKRLLAQFA"
     misc_feature    7393. .7398
                     /gene="SPBC1711.03"
                     /note="gtaaga, splice donor sequence"
     misc_feature    7419. .7431
                     /gene="SPBC1711.03"
                     /note="ttaacgtatctag, splice branch and acceptor"
     misc_feature    7559. .7564
                     /gene="SPBC1711.03"
                     /note="gtaaga, splice donor sequence"
     misc_feature    7590. .7606
                     /gene="SPBC1711.03"
                     /note="ctaacagcaatcttcag, splice branch and acceptor"
     misc_feature    7882. .7887
```

```
                     /gene="SPBC1711.03"
                     /note="gtaagt, splice donor sequence"
     misc_feature    8074. .8086
                     /gene="SPBC1711.03"
                     /note="ctaactatcaaag, splice branch and acceptor"
     misc_feature    8249. .8254
                     /gene="SPBC1711.03"
                     /note="gtatga, splice donor sequence"
     misc_feature    8280. .8293
                     /gene="SPBC1711.03"
                     /note="ttaacatttttttag, splice branch and acceptor"
     gene            join(8867. .8906,9022. .9153,9206. .9996)
                     /gene="SPBC1711.04"
     CDS             join(8867. .8906,9022. .9153,9206. .9996)
                     /gene="SPBC1711.04"
                     /note="SPBC1711.04, len:320, SIMILARITY:Saccharomyces
                     cerevisiae, MTD1_YEAST, methylenetetrahydrofolate
                     dehydrogenase [nad+], (320 aa), fasta scores: opt: 1132,
                     E():0, (58.5% identity in 299 aa)"
                     /codon_start=1
                     /label=SPBC1711.05
                     /product="methylenetetrahydrofolate dehydrogenase [nad+]"
                     /protein_id="CAB88234.1"
                     /db_xref="GI:7630162"
                     /translation="MSNTNKSEPNHSCKVVYASRVAETFVEQLKQHVNLFEFAPKLVG
                     FLSNSDPAARMYADWTNKTCTEIGFQYELREVPKDDLEDAIVEANNDPSVNGIMIYFP
                     VFNDGQDQYLQQVVSPDKDVEGLCHKYVMNMYHNIRHLDPEKTKKSILPCTPLAIVKI
                     LEYLGVYNKIINYGNRLYGKTITIVNRSEIVGRPLAALLANDGAKVYSVDIHNVQCFT
                     RGAGIRSKKHDVADTNFKLEDVAPISDVIICGVPSANYKFPSNLVRDGAVCICFSSEK
                     NFDAASLKEHAGIYVPSIGKVTIAMLLRNLIRLTSYQLNKPVDI"
     misc_feature    8907. .8912
                     /gene="SPBC1711.04"
                     /note="gtaagt, splice donor sequence"
     misc_feature    9005. .9021
                     /gene="SPBC1711.04"
                     /note="ttaattgaatttttacag, splice branch and acceptor"
     misc_feature    9154. .9159
                     /gene="SPBC1711.04"
                     /note="gtaagt, splice donor sequence"
     misc_feature    9187. .9205
                     /gene="SPBC1711.04"
                     /note="ctaatcatgaacaatttag, splice branch and acceptor"
     misc_feature    9553. .9576
                     /gene="SPBC1711.04"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop),
                     [AG].{4}GK[ST], info count = 13.8"
     gene            10273. .11628
                     /gene="SPBC1711.05"
     CDS             10273. .11628
                     /gene="SPBC1711.05"
                     /note="SPBC1711.05, len:452, SIMILARITY:Homo sapiens,
                     O95815, dentin phosphoryn, (788 aa), fasta scores: opt:
                     769, E():8.5e-27, (48.2% identity in 334 aa)"
                     /codon_start=1
                     /label=SPBC1711.05
                     /product="hypothetical serine-rich repeat protein"
                     /protein_id="CAB88235.1"
                     /db_xref="GI:7630163"
                     /translation="MGEISMKSKVCPLIYHFLQENGYVKTAQTFLKETGDKDLAKGKV
                     KKNLLDLLSQKEFLPYLTTEDVGKHKKTKESLEKSNDDSQKISKKGAPPEKAHSSSEA
                     SGSGSSSDESDSSSSESESSSEDNDSSSSSSDSESESSSEDSDSSSSSSDSESESSSE
                     GSDSSSSSSSSSESESSSEDNDSSSSSSDSESESSSEDSDSSSSSSDSESESSSEGSDS
                     SSSSSSSESESSSEDNDSSSSSSDSESESSSEDSDSSSSSSDSESESSSKDSDSSSNS
                     SDSEDDSSSDSSDSESESSSEDSDSTSSSSDSDSSSSSEDGNSNTDTTTSGEVSAQSS
                     TNSTSSEESTSVKDEDSSKIHDKSLKRKHEDDESSTSTKSSRTTKTPFTRVGDPSQWD
                     FASPALRDNSFNFEDDYGTLANRDLIVTRGKGFRQEKNKKKRGSYRGGRINTEVRSFK

  Query Match             8.1%;  Score 149.8;  DB 8;  Length 38000;
  Best Local Similarity   47.9%;  Pred. No. 3.5e-29;
  Matches  597;  Conservative    0;  Mismatches  632;  Indels   18;  Gaps    5;

Qy    276 GCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGA 335
          | |  ||| |   ||    |   | ||  || |    || |     ||||    |  ||

Db  33459 GAACGGTGAAGAAACTACAGATGGAGCAGAAGGGGTGTTGCAAACAAGTGAAGAGGTAGA 33518

Qy    336 TGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGA 395
          |||      |   |  |  ||  |   || |  | |||   ||  |  | | ||||
Db  33519 TGATGAATTGAAAGTATACAATTTGGATACTTACGATGACGATGAGGAGGAAGCTGAGAC 33578

Qy    396 AGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAG 455
          ||  |  || | ||||  |  |        |   |||| |      |   | ||
Db  33579 AGAAGGGCCCGCTGCTATGTTTAGCAACATTCGTGGATTGGCTTACCATGAAAATGGAGA 33638

Qy    456 TAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTT 515
          || |||| | ||       ||  ||   | ||    ||  | |  ||| | || ||  |
Db  33639 AAAAGATCCATATATCACTATTGATCCTCAGGAACAGGACGATTTGGAAAGAGAGGAAAT 33698

Qy    516 CTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAA 575
            ||   || |  || |  ||| |  |   ||  ||  | ||| |  ||  || | |
Db  33699 GCAAATATTACCAACAGACAGCCTTTTGCTAGCTGCGCGCACAGAAGATAACTTGTCCCA 33758

Qy    576 TTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACT 635
          | || ||| |||| ||| | | |  ||| ||   | |||| || || |||||| | ||
Db  33759 CGTTGAAGTGTATGTATATGAGCCAACAGAGGAAAATCTTTACGTGCATCATGATTTTCT 33818

Qy    636 CTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCC---TAGCCCAGATGA 692
          |  || | |||||| || | |||||||  ||| | |      ||   | ||  |
Db  33819 TCTCCCTACCTTTCCTCTGTGTCTCGAATGGCTAGATTATAAGGTGGGTACGTCTGACAA 33878

Qy    693 TTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCT 752
          || |||||||||| |||||| || |   |   ||||  |||||| | |||||  |
Db  33879 CGCTCCTGGAAATTACGTTGCTGTCGGCACGTTTGATCCTGAAATTGAAATATGGGATTT 33938

Qy    753 TGATATAGTGGACTCTTTAGAGC---CAGTCTTCACACTCGGAAGTAAACTTTCAAAAAA 809
          || ||  | ||  |  | | |   ||| | |      | | ||     |  |   ||
Db  33939 AGACATCATTGATGCAGTTTACCCTGCAGCCGTGTTGGGTGCAGGTGCCTCTCAAGTCAA 33998

Qy    810 GAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGA 869
          ||||||||||         ||  |         |        || || ||||||||  ||||
Db  33999 CAAGAAAAAGAAGAAGTCTAAAAAAATAAATGATAGTTATCACACAGATGCTGTTCTTGC 34058

Qy    870 CCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGT 929
          |||||  | |||          |||  |||| |    ||  || || || | | |||  |
Db  34059 TCTTTCTTCGAACCGAAATGCTCACAATCTTTTGGTTTCTGGTTCTGCAGATACCACACT 34118

Qy    930 AATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAA 989
          ||   |||||||| ||  |    | ||   |          || |    ||  | |||||
Db  34119 AAAGTTGTGGGATTTGAGCACATGCAATTGTGTCAAATCTTTCACGTATCATTCTGACAA 34178

Qy    990 GGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGA 1049
          ||        |  | |   || |    |   ||        |  | || ||    |||||
Db  34179 AGTTTCTTGCTTAGATTGGTATTCCAAAGCTCCATCCGTATTATTGTCAGGTAGTTATGA 34238

Qy   1050 TAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATT 1109
          |||  | |      |  |||   ||   | |  || |||    |     |  | | | |
Db  34239 TAAAACTGCAAAAATAGCTGA---CCTTCGCCTTGAAGAAGCTCCATCGAGTTTTCAAGT 34295

Qy   1110 CAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAG 1169
          | | |  |  | || |  ||  | ||| ||||    |           | || ||     |
Db  34296 TACTAGCGATGTTGAAAATGTTGCCTGGGATCAACATAGCGAAAACAACTTTTTCATTGG 34355

Qy   1170 TACAGATGACGGCTTTGTATATAATTTGGATGC---ACGTTCAGATAAGCCAATTTTTAC 1226
          || ||| | ||| | || ||| |||  |||||    |  |       ||| |  |||
Db  34356 CACTGATAATGGCATAGTTTATTATTGCGATGCTAGAAATCTCTCAAAGTCCGTTTGGCA 34415

Qy   1227 ACTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTG 1286
          | |  | || ||  |||     || ||| ||||  | |||    |   |      |
Db  34416 ATTGCAAGCCCATGATGGACCTATTTCTTGTCTCAGTGTTAATCCTTCGGTTCCTTCTTT 34475

Qy   1287 TCTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCC 1346
          | | |  || |  ||  |||| | |   || ||  ||||| ||| |  |   |   ||
Db  34476 TGTAGCCACAGGCTCCACTGATAGAGTGGTTAAACTCTGGAACACTTCCGATAGTTCACC 34535

Qy   1347 AAGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCC 1406
          |   | |||  |||||| ||  || |  | |||  | | ||   ||  || | |    |
Db  34536 GAAGATGGTTGTTTCTAGAGATTTGGATGTTGGACGTGTGTTTACTTGCTCCTTTACTAC 34595
```

```
Qy    1407 TGATTTGCCA------TTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTG 1460
           ||     ||      ||   |  ||       ||    |||| ||   |||| |||||
Db   34596 GGACGAATCAACTGCTTTCCATCTAGCTGCCTCCGGAAGCAAAGGAGTCGTTCGTGTCTG 34655

Qy    1461 GGATATAAGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAG 1507
           ||| |     ||     ||   ||     |||| |||| |||  |||
Db   34656 GGACACTGCAACTAATCCTGGTGTTCGAAAAGCGTTTGAAAGTAGAG 34702
```

Search completed: March 19, 2003, 15:53:05
Job time : 62430 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 18, 2003, 19:15:45 ; Search time 865 Seconds
(without alignments)
4824.223 Million cell updates/sec

Title: US-09-315-355A-47
Perfect score: 1853
Sequence: 1 gatccctgagcgtgtggcag.....aaaagtaattccttaaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1771	95.6	1881	23	AAS70866	DNA encoding novel
	2	1450	78.3	2584	24	ABQ54496	Human ovarian anti
	3	517	27.9	626	21	AAC04021	Human secreted pro
	4	239.8	12.9	596	24	ABQ60025	Human colon cancer
c	5	188.4	10.2	17197	22	AAK79870	Human immune/haema
c	6	150	8.1	150	24	ABN96228	Gene #2726 used to
	7	122.8	6.6	381	21	AAH31017	Human colon cancer
c	8	119.8	6.5	796	24	ABQ89158	Human prostate exp
	9	108.8	5.9	1965	23	ABL11659	Drosophila melanog
	10	108.8	5.9	3965	23	ABL11658	Drosophila melanog
	11	108.8	5.9	4826	23	ABL13260	Drosophila melanog
	12	106	5.7	1566	21	AAC51036	Arabidopsis thalia
	13	104	5.6	553	20	AAX21042	Polynucleotide seq
	14	74	4.0	261	23	AAS70865	DNA encoding novel
	15	69.6	3.8	211	23	ABV17865	Human prostate exp
	16	69.6	3.8	335	23	ABV47656	Human prostate exp
c	17	52.8	2.8	766	22	ABA49279	Human breast cell
c	18	52.8	2.8	766	22	ABA67193	Human foetal liver
c	19	52.8	2.8	766	22	ABA34288	Probe #12754 for g
c	20	52.8	2.8	766	22	AAI22114	Probe #12047 for g
c	21	52.8	2.8	766	22	AAI47409	Probe #16095 used
c	22	52.8	2.8	766	22	AAI07812	Probe #7803 used t
c	23	52.8	2.8	1944	22	ABA44123	Human breast cell
c	24	52.8	2.8	1944	22	ABA54575	Human foetal liver
c	25	52.8	2.8	1944	22	ABA24359	Probe #2825 for ge
c	26	52.8	2.8	1944	22	AAI12874	Probe #2807 for ge
c	27	52.8	2.8	1944	22	AAI34231	Probe #2917 used t
c	28	52.8	2.8	1944	22	AAI02791	Probe #2782 used t
	29	49.2	2.7	489	22	AAH06751	Human cDNA clone (
	30	49	2.6	1959	22	ABA45382	Human breast cell
	31	49	2.6	1959	22	ABA55872	Human foetal liver
	32	49	2.6	1959	22	ABA25546	Probe #4012 for ge
	33	49	2.6	1959	22	AAK04090	Human brain expres
	34	49	2.6	1959	22	AAK29571	Human bone marrow
	35	49	2.6	1959	22	AAI14146	Probe #4079 for ge
	36	49	2.6	1959	22	AAI35531	Probe #4217 used t
	37	49	2.6	1959	22	AAI03994	Probe #3985 used t
	38	49	2.6	1959	24	ABS04127	Human genome-deriv
	39	48.4	2.6	1893	22	AAH14157	Human cDNA sequenc
	40	48.4	2.6	2518	20	AAV71745	Human V3 loop HIV
	41	48.4	2.6	2518	22	AAC92553	Human nucleolin DN
	42	48.4	2.6	3201	23	ABV22156	Human prostate exp
	43	48.4	2.6	3201	23	ABV23046	Human prostate exp
	44	47.8	2.6	583	22	ABA50512	Human breast cell
	45	47.8	2.6	583	22	ABA68463	Human foetal liver

ALIGNMENTS

RESULT 1
AAS70866
ID AAS70866 standard; cDNA; 1881 BP.
XX
AC AAS70866;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6670.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG06679.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 6670; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1881 BP; 530 A; 376 C; 473 G; 502 T; 0 other;

Query Match 95.6%; Score 1771; DB 23; Length 1881;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1843; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

```
Qy       1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCT-GGTTTCT 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      23 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGGTTTCT 82

Qy      60 AGCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGC 119
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      83 AGCGTGACACGGCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGC 142

Qy     120 CTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGA 179
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     143 CTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGGAGG 202

Qy     180 AGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGA 239
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     203 AGTAAAACGCCTCATTGCTGAGGCAGAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGA 262

Qy     240 TGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCA-TGCAGAGTGCACGCACCCAGGCAC 298
          ||||||| |||||||||||||| |||||| || | |||| ||||||||||||||||||||
Db     263 TGAAGAGAAGACAGGCAGTCCTTTAGAAGAAGGGATTGCAAAGTGCACGCACCCAGGCAC 322

Qy     299 G-CCCAAGAGAGCCCCTGGA-GGATGGTGACCCAGAGGA-TGACAGGACGCTTGATGATG 355
          | |||||||||||||||| || |||||||||||||||||| ||||||||||||||||||||
Db     323 GCCCCAAGAGAGCCCCTCGAGGGATGGTGACCCAGAGGATTGACAGGACGCTTGATGATG 382

Qy     356 ATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGA 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     383 ATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGA 442

Qy     416 CTCTTGG-TGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTAC 474
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     443 CTCTTGGTTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTAC 502

Qy     475 GTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     503 GTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGAT 562

Qy     535 AATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTAT 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     563 AATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTAT 622

Qy     595 AATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     623 AATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTG 682

Qy     655 AGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCT 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     683 AGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCT 742

Qy     715 GTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAG 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     743 GTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAG 802

Qy     775 CCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGT 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     803 CCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGT 862

Qy     835 TCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGA 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     863 TCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGA 922

Qy     895 AATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGG 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     923 AATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGG 982

Qy     955 AAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCA 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     983 AAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCA 1042

Qy    1015 TTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1043 TTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGC 1102

Qy    1075 CGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACT 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1103 CGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACT 1162

Qy    1135 TGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAAT 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1163 TGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAAT 1222

Qy    1195 TTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCT 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1223 TTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCT 1282

Qy    1255 GGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATAC 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1283 GGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATAC 1342

Qy    1315 GTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAA 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1343 GTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAA 1402

Qy    1375 ATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGA 1434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1403 ATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGA 1462

Qy    1435 GGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCA 1494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1463 GGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCA 1522

Qy    1495 TTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCT 1554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1523 TTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCT 1582

Qy    1555 TTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTG 1614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1583 TTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTG 1642

Qy    1615 CTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAAC 1674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1643 CTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAAC 1702
```

```
Qy   1675 CATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGT 1734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1703 CATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGT 1762

Qy   1735 GGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTT 1794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1763 GGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTTGTAAGGCTCTT 1822

Qy   1795 GTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATTCCTTAAACAT 1853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1823 GTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATTCCTTAAACAT 1881


RESULT 2
ABQ54496
ID   ABQ54496 standard; cDNA; 2584 BP.
XX
AC   ABQ54496;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HCOQH27 cDNA, SEQ ID NO:376.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 12;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   P-PSDB; ABP41419.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 1; SEQ ID No 376; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents cDNA encoding a human ovarian antigen of the
CC   invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2584 BP; 804 A; 454 C; 529 G; 796 T; 1 other;

  Query Match             78.3%;  Score 1450;  DB 24;  Length 2584;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1450;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    404 CAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATC 463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 CAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATC 188

Qy    464 AAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTA 523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 AAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTA 248

Qy    524 AGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGG 583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 AGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGG 308

Qy    584 TGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTG 643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 TGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTG 368

Qy    644 CATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAA 703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAA 428

Qy    704 ATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGG 763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 ATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGG 488

Qy    764 ACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAG 823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 ACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAG 548

Qy    824 GAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATA 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 GAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATA 608

Qy    884 AGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATA 943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 AGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATA 668

Qy    944 TGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGC 1003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 TGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGC 728

Qy   1004 AGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTT 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 AGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTT 788

Qy   1064 TGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAG 1123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 TGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAG 848

Qy   1124 AGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCT 1183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    849 AGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCT 908

Qy   1184 TTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATG 1243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 TTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATG 968

Qy   1244 ATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAG 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 ATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAG 1028

Qy   1304 CTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTA 1363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 CTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTA 1088

Qy   1364 GGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTT 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 GGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTT 1148

Qy   1424 ATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAG 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 ATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAG 1208

Qy   1484 TAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTA 1543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTA 1268

Qy   1544 TTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCAT 1603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 TTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCAT 1328

Qy   1604 CTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGC 1663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 CTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGC 1388

Qy   1664 GTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCT 1723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 GTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCT 1448

Qy   1724 GCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTT 1783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 GCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTTT 1508

Qy   1784 GTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATT 1843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509 GTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAATT 1568

Qy   1844 CCTTAAACAT 1853
          ||||||||||
Db   1569 CCTTAAACAT 1578
```

RESULT 3
AAC04021
ID AAC04021 standard; cDNA; 626 BP.
XX
AC AAC04021;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 4019.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG04015.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 4019; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 626 BP; 162 A; 142 C; 187 G; 135 T; 0 other;

Query Match 27.9%; Score 517; DB 21; Length 626;
Best Local Similarity 95.2%; Pred. No. 2.3e-138;
Matches 555; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

```
Qy      1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     44 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 103

Qy     61 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    104 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 163

Qy    121 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 223

Qy    181 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 283

Qy    241 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 343

Qy    301 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 403

Qy    361 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    404 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 463

Qy    421 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    464 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 523

Qy    481 CTGAAAGATACA----------------------------GAACAATATGAACGTGAAGA 512
          ||||||||||||                            ||||||||||||||||||||
Db    524 CTGAAAGATACATCAATGACATTTTTTTCCTCTCACTTAGGAACAATATGAACGTGAAGA 583

Qy    513 TTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGA 555
          |||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    584 TTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGA 626


RESULT 4
ABQ60025
ID   ABQ60025 standard; cDNA; 596 BP.
XX
AC   ABQ60025;
XX
DT   02-AUG-2002  (first entry)
XX
DE   Human colon cancer related nucleotide sequence SEQ ID NO:3720.
XX
KW   Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW   genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200229086-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30732.
XX
PR   02-OCT-2000; 2000US-237271P.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Burgess C,  Astle JH,  Carroll E,  Catino TJ,  Dwivedi P,  Molino GA;
PI   Thiaglingam A,  Lewis ME;
XX
DR   WPI; 2002-426115/45.
XX
PT   New isolated nucleic acid that is differentially expressed in cancer
PT   tissues useful for determining the presence of colon cancer in a cell
PT   or tissue type, and in antisense therapy  -
XX
PS   Claim 1; Fig 1; 796pp; English.
XX
CC   ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC   expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC   encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC   used in antisense therapy. An antibody immunoreactive with a polypeptide
CC   encoded by (I) is useful for detecting cancer in a patient sample, and
CC   for detecting the presence or absence of a polynucleotide encoded by a
CC   nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC   from (I) can be used for determining the presence of a nucleic acid which
CC   hybridises to (I), and for determining the phenotype of cells in a sample
CC   of cells from a patient. (I) is useful for determining the presence of
CC   colon cancer in a cell or tissue type, for determining the presence or
CC   state of other type of cancer, in antisense therapy, to generate
CC   macroarrays on a solid surface, to identify a chromosome on which the
CC   corresponding gene resides, and in tissue profiling, forensics, genetic
CC   analysis, mapping and diagnostic applications. (I) can be used to raise
CC   antibodies, and to screen for peptide analogues and antagonists.
XX
SQ   Sequence 596 BP; 179 A; 120 C; 116 G; 158 T; 23 other;

  Query Match             12.9%;  Score 239.8;  DB 24;  Length 596;
  Best Local Similarity   97.2%;  Pred. No. 1.7e-58;
  Matches  241;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    371 ACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTC 430
          |||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      1 ACGACTTANATAAATATGATGAGGAAGGTGACCCAGATGCTGANACTCTTGGTGAATCTC 60

Qy    431 TCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATA 490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATA 120

Qy    491 CAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTG 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTG 180

Qy    551 GCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACT 610
          |||||||||||||||||||| ||||||| | ||| ||||||||||||||||||| ||||||
Db    181 GCCGAGCTGAACAGGACCANTGCAATTNATAGGCGCATGTTTATAATCAAGAANAAGACT 240

Qy    611 CTTTTTAT 618
          ||||||||
Db    241 CTTTTTAT 248


RESULT 5
AAK79870/c
ID   AAK79870 standard; DNA; 17197 BP.
XX
AC   AAK79870;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34682.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
```

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 34682; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 17197 BP; 5296 A; 3779 C; 3387 G; 4735 T; 0 other;

Query Match 10.2%; Score 188.4; DB 22; Length 17197;
Best Local Similarity 99.5%; Pred. No. 7.4e-43;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    219 AGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCA 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14373 AGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCA 14314

Qy    279 GAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGA 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14313 GAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGA 14254

Qy    339 CAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGG 398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14253 CAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGG 14194

Qy    399 TGACCCAGAT 408
          |||||||| |
Db  14193 TGACCCAGGT 14184
```

```
RESULT 6
ABN96228/c
ID   ABN96228 standard; DNA; 150 BP.
XX
AC   ABN96228;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2726 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2726; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 150 BP; 44 A; 33 C; 28 G; 45 T; 0 other;

  Query Match             8.1%;  Score 150;  DB 24;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-33;
  Matches  150;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     660 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 GGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGG 91

Qy     720 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90 AAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGT 31

Qy     780 CTTCACACTCGGAAGTAAACTTTCAAAAAA 809
           ||||||||||||||||||||||||||||||
Db      30 CTTCACACTCGGAAGTAAACTTTCAAAAAA 1
```

```
RESULT 7
AAH31017
ID   AAH31017 standard; cDNA; 381 BP.
XX
AC   AAH31017;
XX
DT   27-JUL-2001  (first entry)
XX
DE   Human colon cancer cell line Km12L4-A cDNA library derived sequence #951.
XX
KW   Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KW   detection; colon cancer cell line Km12L4-A; ss.
XX
OS   Homo sapiens.
XX
PN   WO200018916-A2.
XX
PD   06-APR-2000.
XX
PF   23-SEP-1999;   99WO-US22226.
XX
PR   28-SEP-1998;  98US-0102161.
PR   28-SEP-1998;   98US-0102180.
PR   29-SEP-1998;   98US-0102380.
PR   08-OCT-1998;   98US-0103815.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-293155/25.
XX
PT   Polynucleotide library comprising 1079 defined sequences,  useful in
PT   the form of an array to detect cancer or susceptibility to cancer -
XX
PS   Claim 1; Page 455; 502pp; English.
XX
CC   The present invention describes a library of polynucleotides comprising
CC   1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC   are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC   one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC   (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC   specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC   of detecting differentially expressed genes correlated with a cancerous
CC   state of a mammalian cell comprising detecting a gene product encoded by
CC   65 of the 1079 sequences given in the specification. The polynucleotides
CC   are used to monitor patients having (or susceptible) to cancer to detect
CC   potentially malignant events at a molecular level before they are
CC   detectable at a gross morphological level. The polynucleotides are also
CC   useful for monitoring the efficacy of various therapies and preventive
CC   interventions. Polynucleotide probes based on the disclosed sequences
CC   are useful for chromosome mapping and detection of transcription levels.
CC   The 1079 polynucleotide sequences were derived from a human colon cancer
CC   cell line Km12L4-A cDNA library.
XX
SQ   Sequence 381 BP; 67 A; 115 C; 129 G; 70 T; 0 other;

  Query Match             6.6%;  Score 122.8;  DB 21;  Length 381;
  Best Local Similarity   94.8%;  Pred. No. 6.6e-25;
  Matches  127;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      38 CTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGCA 97
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCCCTATGCAGTCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGCA 60

Qy      98 GCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACA 157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      61 GCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACA 120

Qy     158 AGGTAGAGCTGAGT 171
           ||||  ||  ||
Db     121 AGGTGAGGCCTGGT 134


RESULT 8
ABQ89158/c
ID   ABQ89158 standard; cDNA; 796 BP.
XX
AC   ABQ89158;
XX
DT   27-SEP-2002  (first entry)
XX
DE   Human prostate expressed polynucleotide SEQ ID NO 414.
XX
KW   Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200255700-A2.
XX
PD   18-JUL-2002.
XX
PF   07-DEC-2001; 2001WO-US47349.
XX
PR   07-DEC-2000; 2000US-254648P.
PR   13-MAR-2001; 2001US-275688P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Escobedo J,  Garcia PD,  Kassam A,  Lamson G,  Drmanac R;
PI   Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I,  Leshkowitz D;
PI   Kita D,  Garcia V,  Jones WL,  Stache-Crain B,  Scott EM;
XX
DR   WPI; 2002-557824/59.
XX
PT   New genes and gene products isolated from human prostate, useful for
PT   treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
PT   cancer), or as vaccines for treating or preventing these diseases  -
XX
PS   Claim 1; SEQ ID NO 414; 186pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any of
CC   1477 sequences or its fragment, degenerate variant, antisense or
CC   complement. The polynucleotides and gene products are useful for treating
CC   or diagnosing tumour or cancer (e.g. prostate cancer, breast cancer,
CC   lung cancer or medullary carcinoma) in a subject (e.g. cattle, dogs,
CC   cats, rabbits, horse or human). The polynucleotides and polypeptides are
CC   also useful as vaccines for treating or preventing these diseases. The
CC   polynucleotides are useful for gene therapy. The present sequence is that
CC   of one of a group of polynucleotides (ABQ88745-ABQ90015) disclosed
CC   electronically as sequences of the invention. However only 1271
CC   polynucleotide sequences are given, whereas 1477 polynucleotides and 91
CC   proteins are claimed.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequence.
XX
SQ   Sequence 796 BP; 254 A; 161 C; 170 G; 178 T; 33 other;

  Query Match             6.5%;  Score 119.8;  DB 24;  Length 796;
  Best Local Similarity   94.7%;  Pred. No. 7.4e-24;
  Matches  124;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     589 GTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATAT 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 GTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATAT 248

Qy     649 CCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTAC 708
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     247 CCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGTAATTAG 188

Qy     709 ATTGCTGTAGG 719
           |   ||  |||
Db     187 AAAACTTAAGG 177


RESULT 9
ABL11659
ID   ABL11659 standard; cDNA; 1965 BP.
XX
AC   ABL11659;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 29459.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB67556.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 29459; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1965 BP; 530 A; 486 C; 510 G; 439 T; 0 other;

  Query Match             5.9%;  Score 108.8;  DB 23;  Length 1965;
  Best Local Similarity   48.0%;  Pred. No. 1.8e-20;
  Matches  462;  Conservative    0;  Mismatches  467;  Indels   33;  Gaps    4;

Qy     493 GAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGC 552
           ||| | |  ||    || ||     |||| |||||||| || |||| ||  |    ||
Db     425 GAAGACTCCGAGGCCGAGGACGAGGTGATCAAGCCCAGCGACAACCTCATTCTAGTGGGT 484

Qy     553 CGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCT 612
           | | | || | |||     |   | ||||||   |||| || |||| || || ||  ||
Db     485 CACGTTCAAGACGACGCCGCCTCCATGGAGGTGTGGGTTTTCAACCAGGAGGAGGAGGCT 544

Qy     613 TTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAAT 672
           | ||    ||||| ||  | ||  || |    | |||||||  | | | || ||| |||||
```

```
Db    545 CTCTACACCCACCACGACTTTCTGCTGCCAAGCTTTCCTCTGTGCATCGAGTGGATGAAT 604

Qy    673 TTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCT 732

Db    605 CACGACGCGGGCAGCGAAAA---GGCGGGCAACATGTGCGCCATCGGCTGCATGGATCCG 661

Qy    733 GTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGA 792

Db    662 ATAATCACAGTCTGGGATCTAGACATACAGGACGCTATCGAGCCCACATTTAAGCTGGGT 721

Qy    793 AGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCAT 852

Db    722 TCCAAAGGCAGCCGGAAGCAGAACAAAGA------------------GCAGTATGGACAC 763

Qy    853 ACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCA 912

Db    764 AAGGACGCCGTGCTGGATCTCTCTTGGAACACCAACTTTGAGCACATTCTGGCCAGCGGG 823

Qy    913 TCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTC 972

Db    824 TCCGTGGACCAAACTGTGATTCTGTGGGACATGGACGAGGGCCAGCCTCATACAACCATT 883

Qy    973 GCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTG 1032

Db    884 ACCGCTTTTGGCAAACAGATTCAGTCGCTGGAATTCCATCCGCAAGAGGCTCAAAGCATT 943

Qy   1033 ATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGC 1092

Db    944 CTTACCGGCTGTGCCGATGGATACGTGCGACTCTTCGATTGCCGCGACGCTGAGGGCGTC 1003

Qy   1093 CA---------TCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCAC 1143

Db   1004 AACTCGTCCAGCATTGAGTGGAAAGTTGACGGTGAAGTGGAGAAGGTCCTGTGGCATCCC 1063

Qy   1144 TTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCA 1203

Db   1064 ACACAGACCGACTACTTCATCGTGGGCACCAACGATGGCACCTTGCATTACGCCGACAAA 1123

Qy   1204 CGTTC---AGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260

Db   1124 CGTTCTCCTGGACAACTGCTGTGGTCCGTAAAGGCCCACAACGAGGAAATCTCCGGTGTG 1183

Qy   1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320

Db   1184 TGCTTCAACAACCAGAAGCCTAATCTGCTGACCTCCACCTCCACGGAGGGCACCCTAAAG 1243

Qy   1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380

Db   1244 GTGTGGAACTTTGATGGCACAGAGGCAAAGCACGTCTACGAGCACGAGTTCAACATGGGT 1303

Qy   1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440

Db   1304 CGCTTGCAGTGCATGCGCCAGTGCCCCGAGGATCCCTACACCCTGGCCTTCGGCGGAGAG 1363

Qy   1441 AA 1442

Db   1364 AA 1365
```

RESULT 10
ABL11658
ID ABL11658 standard; cDNA; 3965 BP.
XX
AC ABL11658;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29456.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB67555.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 29456; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3965 BP; 1088 A; 946 C; 979 G; 952 T; 0 other;

Query Match 5.9%; Score 108.8; DB 23; Length 3965;
Best Local Similarity 48.0%; Pred. No. 2.7e-20;
Matches 462; Conservative 0; Mismatches 467; Indels 33; Gaps 4;

```
Qy    493 GAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGC 552

Db   1425 GAAGACTCCGAGGCCGAGGACGAGGTGATCAAGCCCAGCGACAACCTCATTCTAGTGGGT 1484

Qy    553 CGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCT 612

Db   1485 CACGTTCAAGACGACGCCGCCTCCATGGAGGTGTGGGTTTTCAACCAGGAGGAGGAGGCT 1544

Qy    613 TTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAAT 672

Db   1545 CTCTACACCCACCACGACTTTCTGCTGCCAAGCTTTCCTCTGTGCATCGAGTGGATGAAT 1604

Qy    673 TTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCT 732

Db   1605 CACGACGCGGGCAGCGAAAA---GGCGGGCAACATGTGCGCCATCGGCTGCATGGATCCG 1661

Qy    733 GTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGA 792

Db   1662 ATAATCACAGTCTGGGATCTAGACATACAGGACGCTATCGAGCCCACATTTAAGCTGGGT 1721

Qy    793 AGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCAT 852

Db   1722 TCCAAAGGCAGCCGGAAGCAGAACAAAGA------------------GCAGTATGGACAC 1763

Qy    853 ACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCA 912

Db   1764 AAGGACGCCGTGCTGGATCTCTCTTGGAACACCAACTTTGAGCACATTCTGGCCAGCGGG 1823

Qy    913 TCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTC 972

Db   1824 TCCGTGGACCAAACTGTGATTCTGTGGGACATGGACGAGGGCCAGCCTCATACAACCATT 1883
```

```
Qy    973 GCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTG 1032
          | |          |  || | ||  | ||| | || |||||    || || || |   |
Db   1884 ACCGCTTTTGGCAAACAGATTCAGTCGCTGGAATTCCATCCGCAAGAGGCTCAAAGCATT 1943

Qy   1033 ATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGC 1092
          || | ||||     |||    |  |||     | |  || |||||      | || |    |
Db   1944 CTTACCGGCTGTGCCGATGGATACGTGCGACTCTTCGATTGCCGCGACGCTGAGGGCGTC 2003

Qy   1093 CA---------TCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCAC 1143
            |                  |||| | |     ||  | | ||||  ||    ||| ||| |
Db   2004 AACTCGTCCAGCATTGAGTGGAAAGTTGACGGTGAAGTGGAGAAGGTCCTGTGGCATCCC 2063

Qy   1144 TTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCA 1203
                |     | ||| | |   | ||  | || |||    |  || |    ||   |
Db   2064 ACACAGACCGACTACTTCATCGTGGGCACCAACGATGGCACCTTGCATTACGCCGACAAA 2123

Qy   1204 CGTTC---AGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260
          |||||    |   | |   | |   | | || || ||||| || ||||||||| ||| |
Db   2124 CGTTCTCCTGGACAACTGCTGTGGTCCGTAAAGGCCCACAACGAGGAAATCTCCGGTGTG 2183

Qy   1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320
            | | ||  || |  |           ||     | | ||  | ||       | | |||
Db   2184 TGCTTCAACAACCAGAAGCCTAATCTGCTGACCTCCACCTCCACGGAGGGCACCCTAAAG 2243

Qy   1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380
          | ||| || |    ||      | ||| |  ||  |        ||  | || |||||
Db   2244 GTGTGGAACTTTGATGGCACAGAGGCAAAGCACGTCTACGAGCACGAGTTCAACATGGGT 2303

Qy   1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440
            |   ||         |||||  ||    ||  |   |     |||||  || ||  |
Db   2304 CGCTTGCAGTGCATGCGCCAGTGCCCCGAGGATCCCTACACCCTGGCCTTCGGCGGAGAG 2363

Qy   1441 AA 1442
          ||
Db   2364 AA 2365
```

RESULT 11
ABL13260
ID ABL13260 standard; cDNA; 4826 BP.
XX
AC ABL13260;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34262.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69157.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 34262; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4826 BP; 1258 A; 1202 C; 1196 G; 1170 T; 0 other;

Query Match 5.9%; Score 108.8; DB 23; Length 4826;
Best Local Similarity 48.0%; Pred. No. 3.1e-20;
Matches 462; Conservative 0; Mismatches 467; Indels 33; Gaps 4;

```
Qy    493 GAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGC 552
          |||  |  |  ||    || ||     |||| |||||||| || || || ||  |   ||
Db    369 GAAGACTCCGAGGCCGAGGACGAGGTGATCAAGCCCAGCGACAACCTCATTCTAGTGGGT 428

Qy    553 CGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCT 612
          |  | | || | |||      |    | ||||||  ||||  || || || || ||  ||
Db    429 CACGTTCAAGACGACGCCGCCTCCATGGAGGTGTGGGTTTTCAACCAGGAGGAGGAGGCT 488

Qy    613 TTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAAT 672
          | ||    ||||| ||  | ||  || |    | |||||| | || || ||| |||||
Db    489 CTCTACACCCACCACGACTTTCTGCTGCCAAGCTTTCCTCTGTGCATCGAGTGGATGAAT 548

Qy    673 TTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCT 732
            ||  |  ||   ||  |    | || ||       ||  | ||   ||||   ||
Db    549 CACGACGCGGGCAGCGAAAA---GGCGGGCAACATGTGCGCCATCGGCTGCATGGATCCG 605

Qy    733 GTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGA 792
          | ||   ||| ||||| || || ||| |||| || | |||||   || |  || ||
Db    606 ATAATCACAGTCTGGGATCTAGACATACAGGACGCTATCGAGCCCACATTTAAGCTGGGT 665

Qy    793 AGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCAT 852
            |||          ||| |||| ||  |                    | || ||
Db    666 TCCAAAGGCAGCCGGAAGCAGAACAAAGA----------------GCAGTATGGACAC 707

Qy    853 ACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCA 912
          |  || || || || || || || ||||| |       |  ||| | || || |
Db    708 AAGGACGCCGTGCTGGATCTCTCTTGGAACACCAACTTTGAGCACATTCTGGCCAGCGGG 767

Qy    913 TCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTC 972
          || |  ||| | |||| |||||||||| ||| |  ||| |  ||     | | | |
Db    768 TCCGTGGACCAAACTGTGATTCTGTGGGACATGGACGAGGGCCAGCCTCATACAACCATT 827

Qy    973 GCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTG 1032
          | |          |  || | ||  | ||| | || |||||    || || || |   |
Db    828 ACCGCTTTTGGCAAACAGATTCAGTCGCTGGAATTCCATCCGCAAGAGGCTCAAAGCATT 887

Qy   1033 ATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGC 1092
          || | ||||     |||    |  |||     | |  || |||||      | || |    |
Db    888 CTTACCGGCTGTGCCGATGGATACGTGCGACTCTTCGATTGCCGCGACGCTGAGGGCGTC 947

Qy   1093 CA---------TCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCAC 1143
            |                  |||| | |     ||  | | ||||  ||    ||| ||| |
Db    948 AACTCGTCCAGCATTGAGTGGAAAGTTGACGGTGAAGTGGAGAAGGTCCTGTGGCATCCC 1007

Qy   1144 TTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCA 1203
                |     | ||| | |   | ||  | || |||    |  || |    ||   |
Db   1008 ACACAGACCGACTACTTCATCGTGGGCACCAACGATGGCACCTTGCATTACGCCGACAAA 1067

Qy   1204 CGTTC---AGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGGTCTT 1260
          |||||    |   | |   | |   | | || || ||||| || ||||||||| ||| |
Db   1068 CGTTCTCCTGGACAACTGCTGTGGTCCGTAAAGGCCCACAACGAGGAAATCTCCGGTGTG 1127

Qy   1261 GATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGTGAAG 1320
            | | ||  || |           ||     | | ||  | ||       | | |||
```

```
Db   1128 TGCTTCAACAACCAGAAGCCTAATCTGCTGACCTCCACCTCCACGGAGGGCACCCTAAAG 1187

Qy   1321 ATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAAATGGGA 1380
          | ||| || |    ||       | |||  |  ||  |        ||  | || |||||
Db   1188 GTGTGGAACTTTGATGGCACAGAGGCAAAGCACGTCTACGAGCACGAGTTCAACATGGGT 1247

Qy   1381 GTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGAGGTCAA 1440
            |    ||          |||||  ||    ||  |  |     |||||  ||  ||  |
Db   1248 CGCTTGCAGTGCATGCGCCAGTGCCCCGAGGATCCCTACACCCTGGCCTTCGGCGGAGAG 1307

Qy   1441 AA 1442
          ||
Db   1308 AA 1309
```

RESULT 12
AAC51036
ID AAC51036 standard; DNA; 1566 BP.
XX
AC AAC51036;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67037.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.7%; Score 106; DB 21; Length 1566;

Best Local Similarity 48.2%; Pred. No. 1e-19;
Matches 378; Conservative 0; Mismatches 385; Indels 21; Gaps 2;

```
Qy    460 GATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTG 519
          |||| | |||   | | |  |    | ||    ||  || | ||| | ||| ||
Db    460 GATCCATATCTGAAAGATGTTGATGATGAAGATGATGAAGAGGATATTGATGACACGACA 519

Qy    520 ATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTA 579
          | ||||| |  |||    | ||  ||||||   |   |||| | |||    || || ||
Db    520 GTCAAGCCGACCGATTCGGTGATTATTTGTGCTAGGAATGAAGATGATGTTAGCCATCTA 579

Qy    580 GAGGTGCATGTTTATAATCAAGAAGAAGACTCTT------TTTATGTACACCATGATATA 633
          |||||  || | ||  |  ||       | ||||       | |||||| || ||| | |||
Db    580 GAGGTTTATTTATACGAGGAATCGTCTGGCTCTCCAAACATGTATGTTCATCATCACATA 639

Qy    634 CTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGAT 693
          |  |  | | ||  ||  || ||   | ||||||  ||| |  ||| ||    ||  ||
Db    640 ATTATACCAGAATTCCCATTGTGTACTGCATGGCTTGATTGTCCTCTTAAAGGAGGGGAA 699

Qy    694 TCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTT 753
          |     |  ||   |||  || ||   || |||  | ||||| |||
Db    700 AAAGGGAATTTTGTAGCTATTGGTTCAAAGGATACACCGACAATCGAGATATGGGATCTT 759

Qy    754 GATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACT------------- 800
          ||  |  ||||    |   ||    |  ||| ||| | | ||
Db    760 GACGTTAGGGACGAGGTGCTACCTTGTATACAACTAGGAGGAATAGAGGAGATGATAGTG 819

Qy    801 --TTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGAT 858
            |  |||||||||| ||||||   ||  ||  | |  || |||   | |||||  ||
Db    820 AGTAAGAAAAAGAAGAGCAAGAAACAGAAACCGAAATTCAAAGAAGGTAGCCATACTGAG 879

Qy    859 GCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCT 918
          | || ||||  ||| | ||||| |||     | || |  | ||| |||||    ||
Db    880 TCAGTACTTGGTCTTGCTTGGAACAAGGAGTTTCGGAACATACTTGCTAGTGCTAGTGCC 939

Qy    919 GACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTA 978
          ||||| |  || |  | ||||||  || |    ||||          ||   | |
Db    940 GACAAAAAAGTTAAGGTCTGGGATGTGGCTACTGGGACGTGTAAGATTACTATGGAGCAT 999

Qy    979 CACACAGACAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCT 1038
          |||||| |  |||| ||| |  |     |  | |  | ||  || |    ||  |   |
Db   1000 CACACAAAGGAGGTTCAAGCGGTTGCTTGGAACCATTATGCTCCAGAAGTGCTTCTCAGT 1059

Qy   1039 GGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGA 1098
          || || | |||| |  | || |   || | ||| |  ||  ||  |     |     |
Db   1060 GGGTCGTTTGATCAAACTGTTGTCATGAAAGACGGAAGACAACCTTCACACTCGGGTTTC 1119

Qy   1099 ATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCAT 1158
          |  |||    |||       |  | || ||  |  | ||| |||  |  ||
Db   1120 AAATGGTCTGTCATGTCTGATGTCGAAAGCTTAGCCTGGGATCCTCATTGCGAACACTCC 1179

Qy   1159 TTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCA 1218
          ||  |||  |||   || || ||    ||  |   ||| |||  |||| | |   || ||
Db   1180 TTTGTGGTAAGTCTTGAAGATGGAACCGTGAAGGGTTTTGATATACGTGCTGCACAGTCA 1239

Qy   1219 ATTT 1222
            ||
Db   1240 GGTT 1243
```

RESULT 13
AAX21042
ID AAX21042 standard; DNA; 553 BP.
XX
AC AAX21042;
XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX

```
OS   Treponema pallidum.
XX
PN   WO9859034-A2.
XX
PD   30-DEC-1998.
XX
PF   23-JUN-1998;   98WO-US13041.
XX
PR   24-JUN-1997;   97US-0050667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Fraser CM;
XX
DR   WPI; 1999-081273/07.
XX
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
XX
PS   Claim 1; Page 1034-1035; 1150pp; English.
XX
CC   AAX20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
XX
SQ   Sequence 553 BP; 144 A; 105 C; 150 G; 153 T; 1 other;

  Query Match             5.6%;  Score 104;  DB 20;  Length 553;
  Best Local Similarity   80.3%;  Pred. No. 2.1e-19;
  Matches  122;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy   1238 ACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTG 1297
           ||| ||      | |  || |||||||| |||||||| |||||||||||||| |||||||
Db     34 ACACTGTCTTCCTTTTAGGGCTTGATCTCAGCAGTCAGATCAAGGGCTGTCTTGTGACTG 93

Qy   1298 CTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTC 1357
           | ||||| |||||||||||||||||||||||||  | ||||| ||||| ||||| ||||
Db     94 CATCAGCAGACAAATACGTGAAGATCTGGGACATACTGGGAGACAGGCCGAGTCTGGTTC 153

Qy   1358 ATTCTAGGGACATGAAAATGGGAGTTCTCTTC 1389
           ||||  |||||||||||||| |    ||| |
Db    154 ATTCCCGGGACATGAAAATGGTAAGAATCTCC 185


RESULT 14
AAS70865
ID   AAS70865 standard; cDNA; 261 BP.
XX
AC   AAS70865;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6669.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06678.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 6669; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 261 BP; 63 A; 73 C; 75 G; 50 T; 0 other;

  Query Match             4.0%;  Score 74;  DB 23;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-11;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     88 ATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAG 147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAG 60

Qy    148 ACACCAGACAAGGT 161
           ||||||||||||||
Db     61 ACACCAGACAAGGT 74


RESULT 15
ABV17865
ID   ABV17865 standard; cDNA; 211 BP.
XX
AC   ABV17865;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 17856.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
```

PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 2951; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 211 BP; 74 A; 41 C; 40 G; 44 T; 12 other;

Query Match 3.8%; Score 69.6; DB 23; Length 211;
Best Local Similarity 84.8%; Pred. No. 1e-09;
Matches 78; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy    805 AAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTC 864
          |||||  |   |||  ||  |    ||||||||||||||||||||||||||||||||||||
Db    112 AAAAAATAATTAAAACAATTATTACAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTC 171

Qy    865 CTTGACCTTTCATGGAATAAGCTAATCAGAAA 896
          |||||||||||||||||||||||||||||| ||
Db    172 CTTGACCTTTCATGGAATAAGCTAATCAGGAA 203
```

Search completed: March 18, 2003, 22:32:43
Job time : 896 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: March 18, 2003, 19:24:19 ; Search time 9898 Seconds
(without alignments)
3031.950 Million cell updates/sec

Title: US-09-315-355A-47
Perfect score: 1853
Sequence: 1 gatccctgagcgtgtggcag.....aaaagtaattccttaaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1148.6	62.0	2539	11	AK009972	AK009972 Mus muscu
c	2	966.6	52.2	1040	9	AL515984	AL515984 AL515984
	3	953.4	51.5	995	9	AL561661	AL561661 AL561661
c	4	950.6	51.3	1052	9	AL582812	AL582812 AL582812
	5	938	50.6	983	9	AL515985	AL515985 AL515985
	6	932	50.3	1073	13	BM547947	BM547947 AGENCOURT
	7	917	49.5	942	9	AL549213	AL549213 AL549213
	8	906	48.9	906	14	BQ228676	BQ228676 AGENCOURT
c	9	880	47.5	902	9	AL576190	AL576190 AL576190
	10	868	46.8	1007	13	BM552364	BM552364 AGENCOURT
	11	865	46.7	894	14	BQ948509	BQ948509 AGENCOURT
	12	847.4	45.7	869	14	BQ224661	BQ224661 AGENCOURT
	13	842.4	45.5	953	14	BQ643631	BQ643631 AGENCOURT
	14	825.4	44.5	941	14	BQ646683	BQ646683 AGENCOURT
	15	823.2	44.4	965	14	BQ651718	BQ651718 AGENCOURT
	16	821.8	44.3	1113	14	BQ064709	BQ064709 AGENCOURT
	17	820.8	44.3	860	11	BC032127	BC032127 Homo sapi
	18	818.8	44.2	866	9	AU125142	AU125142 AU125142
	19	818.2	44.2	1002	13	BM562222	BM562222 AGENCOURT
	20	813.4	43.9	1248	11	AK011891	AK011891 Mus muscu
	21	813.2	43.9	980	14	BM810786	BM810786 AGENCOURT
	22	805.8	43.5	922	14	BQ881882	BQ881882 AGENCOURT
c	23	802.8	43.3	809	9	AL552984	AL552984 AL552984
	24	797.8	43.1	806	14	BQ425859	BQ425859 AGENCOURT
	25	794	42.8	875	12	BG575532	BG575532 602598947
	26	791.6	42.7	1022	13	BM562256	BM562256 AGENCOURT
	27	786.6	42.5	1105	14	BQ276324	BQ276324 AGENCOURT
c	28	783.8	42.3	817	14	BM980858	BM980858 UI-CF-EN1
	29	782.8	42.2	887	11	BC010921	BC010921 Homo sapi
	30	780.8	42.1	1079	14	BM913142	BM913142 AGENCOURT
	31	778.2	42.0	783	13	BM559354	BM559354 AGENCOURT
	32	778.2	42.0	783	14	BM809231	BM809231 AGENCOURT
	33	778.2	42.0	783	14	BM913030	BM913030 AGENCOURT
	34	778.2	42.0	1043	14	BQ070859	BQ070859 AGENCOURT
	35	776.2	41.9	782	13	BM556169	BM556169 AGENCOURT
	36	773.8	41.8	916	14	BQ647936	BQ647936 AGENCOURT
c	37	773.2	41.7	789	9	AA639765	AA639765 nq84a04.s
	38	766.2	41.3	796	13	BM423756	BM423756 AGENCOURT
	39	751.8	40.6	919	14	BQ652250	BQ652250 AGENCOURT
	40	750.8	40.5	772	13	BM452846	BM452846 AGENCOURT
	41	750	40.5	783	9	AU125056	AU125056 AU125056
	42	747.2	40.3	786	14	BM912667	BM912667 AGENCOURT
c	43	747	40.3	777	14	BM978566	BM978566 UI-CF-DU1
	44	747	40.3	983	12	BE796924	BE796924 601587284
c	45	745.8	40.2	783	9	AI281923	AI281923 qm26a08.x

ALIGNMENTS

RESULT 1
AK009972
LOCUS AK009972 2539 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310058A11:homolog to PERIODIC TRYPTOPHAN PROTEIN 1 HOMOLOG (KERATINOCYTE PROTEIN IEF SSP 9502), full insert sequence.
ACCESSION AK009972
VERSION AK009972.1 GI:12845101
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310058A11.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

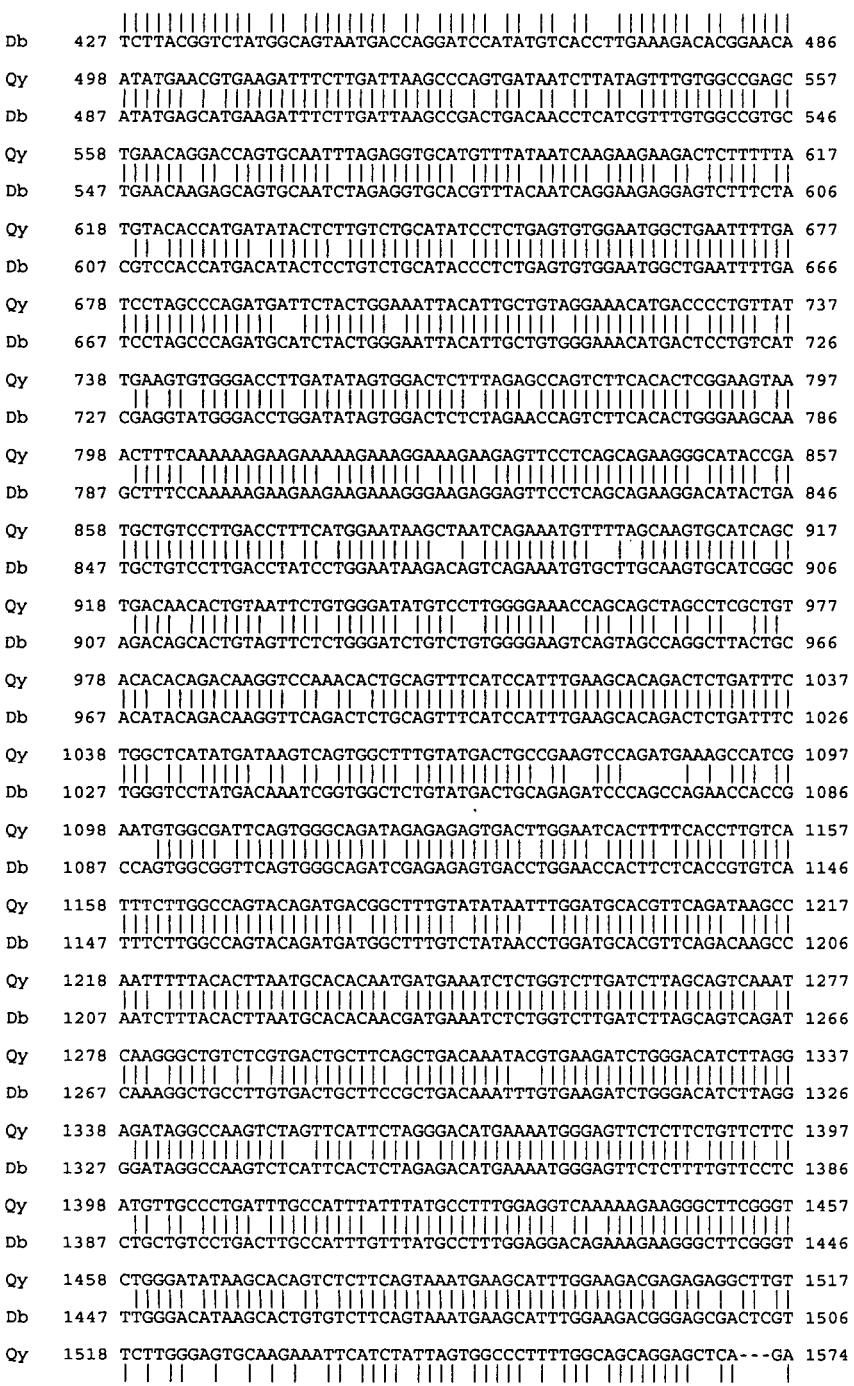
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 2539)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

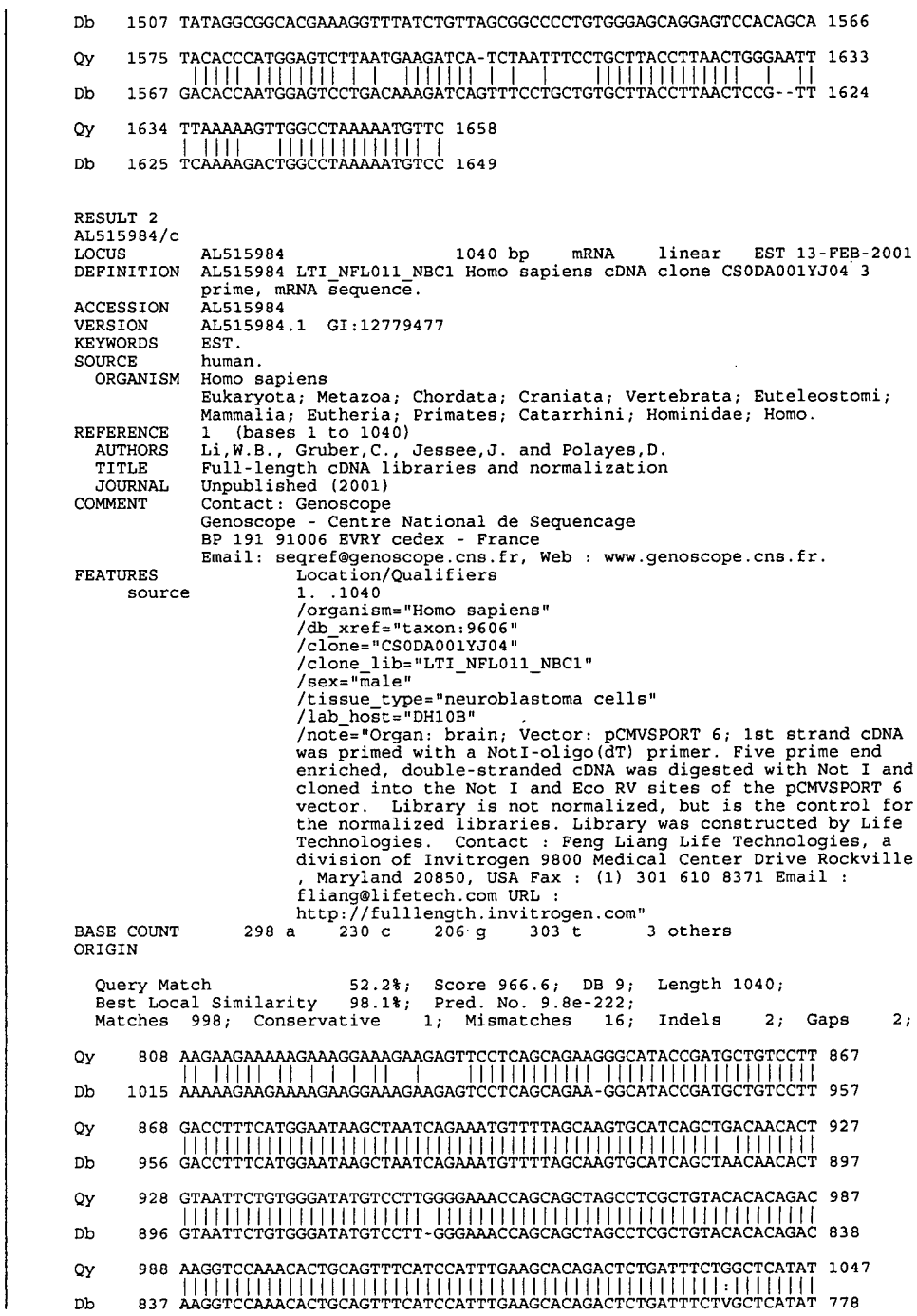
FEATURES Location/Qualifiers
source 1. .2539
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2310058A11"
/db_xref="MGD:MGI:1896893"
/db_xref="taxon:10090"
/clone="2310058A11"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 80. .1585
/note="data source:SPTR, source key:Q13610, evidence:ISS homolog to PERIODIC TRYPTOPHAN PROTEIN 1 HOMOLOG (KERATINOCYTE PROTEIN IEF SSP 9502) putative"
/codon_start=1
/protein_id="BAB26622.1"
/db_xref="GI:12845102"
/db_xref="MGD:MGI:1914735"
/translation="MNRSRQVTCVAWVRRGVAKETPDKVELSKEEVNRLIAEAKGKLQ EEGGSEEEEAGNPSEDGMQSGPTQAPPRESLEDGDPQDDRTLDDDELAGYDLDNYDEE DNPDAETIGESLLGLTVYGSNDQDPYVTLKDTEQYEHEDFLIKPTDNLIVCGRAEQEQ CNLEVHVYNQEEESFYVHHDILLSAYPLSVEWLNFDPSPDASTGNYIAVGNMTPVIEV WDLDIVDSLEPVFTLGSKLSKKKKKKGKRSSSAEGHTDAVLDLSWNKTVRNVLASASA DSTVVLWDLSVGKSVARLTAHTDKVQTLQFHPFEAQTLISGSYDKSVALYDCRDPSQN HRQWRFSGQIERVTWNHFSPCHFLASTDDGFVYNLDARSDKPIFTLNAHNDEISGLDL SSQIKGCLVTASADKFVKIWDILGDRPSLIHSRDMKMGVLFCSSCCPDLPFVYAFGGQ KEGLRVWDISTVSSVNEAFGRRERLVIGGTKGLSVSGPCGSRSPQQTPMES"
polyA_signal 2516. .2521
/note="putative"
polyA_site 2539
/note="putative"

BASE COUNT 686 a 576 c 605 g 672 t
ORIGIN

Query Match 62.0%; Score 1148.6; DB 11; Length 2539;
Best Local Similarity 82.8%; Pred. No. 1.8e-265;
Matches 1362; Conservative 0; Mismatches 274; Indels 9; Gaps 4;

```
Qy     18 CAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGA 77
          | || | ||     | |      || |   ||  || ||  || | | ||    ||   | |
Db     10 CCGCGGAGCTCAAGAGTAGTGTCTCCCAGCTCCCAGGGCTCCAACCTGCGCCGGGTTCAA 69

Qy     78 CTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGT 137
          || || ||||||||||||||||||| ||||| |||||||||||||||||||| |||||||
Db     70 ACCGAAGATCATGAACCGCAGCCGCCAAGTGACCTGCGTGGCCTGGGTCCGCCGCGGCGT 129

Qy    138 GGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGC 197
          |||||||||||| |||||||||||||||||| |||||||| |||||| || || ||||| ||
Db    130 GGCCAAAGAGACTCCAGACAAGGTAGAGCTAAGTAAAGAGGAAGTGAACCGTCTCATCGC 189

Qy    198 TGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAG 257
          ||| ||||| |  || |||||||||||    ||||| ||||| |||||||||| ||||||
Db    190 TGAAGCAAAAGGAAAGTTGCAAGAAGAA---GGTGGGAGTGAAGAAGAGGAGGCAGGCAA 246

Qy    258 TCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGA 317
          |||||||||||||||||||||||||  |  |||||||||  |||| ||| |||||||
Db    247 CCCTTCAGAAGATGGCATGCAGAGTGGCCCGACCCAGGCACCGCCTAGGGAGTCCCTGGA 306

Qy    318 GGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTT 377
          |||||| || ||| | ||||||| | || || || ||||| |||||||| | |||||||||
Db    307 GGATGGGGATCCACAAGATGACCGCACACTGGACGATGACGAGCTGGCGGGGTACGACTT 366

Qy    378 AGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGG 437
          |||||| ||||| || || | | |||||||||||||||  |||||| ||||| |||||
Db    367 AGATAATTATGACGAAGAGGATAACCCAGATGCTGAGACCATTGGCGAGTCTCTGTTGGG 426

Qy    438 TCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACA 497
```





RESULT 2
AL515984/c
LOCUS AL515984 1040 bp mRNA linear EST 13-FEB-2001
DEFINITION AL515984 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA001YJ04 3 prime, mRNA sequence.
ACCESSION AL515984
VERSION AL515984.1 GI:12779477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1040)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .1040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DA001YJ04"
/clone_lib="LTI_NFL011_NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 298 a 230 c 206 g 303 t 3 others
ORIGIN

Query Match 52.2%; Score 966.6; DB 9; Length 1040;
Best Local Similarity 98.1%; Pred. No. 9.8e-222;
Matches 998; Conservative 1; Mismatches 16; Indels 2; Gaps 2;

```
Qy   1048 GATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGA 1107
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    777 GATAAGTCAGTGGCTTTGTATAACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGA 718

Qy   1108 TTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCC 1167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    717 TTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCC 658

Qy   1168 AGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACA 1227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 AGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACA 598

Qy   1228 CTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGT 1287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 CTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGT 538

Qy   1288 CTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCA 1347
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 CTCGTAACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCA 478

Qy   1348 AGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCT 1407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 AGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCT 418

Qy   1408 GATTTGCCATTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATA 1467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 GATTTGCCATTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATA 358

Qy   1468 AGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGT 1527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 AGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGT 298

Qy   1528 GCAAGAAATTCATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAG 1587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 GCAAGAAATTCATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAG 238

Qy   1588 TCTTAATGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCC 1647
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 TCTTAATGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCC 178

Qy   1648 TAAAAATGTTCCATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTG 1707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 TAAAAATGTTCCATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTY 118

Qy   1708 ACTGACATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTC 1767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117 ACTGACATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTC 58

Qy   1768 TTCAGATGGATGGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAA 1824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57 TTCAGATGGATGGTTTGTAAGGCTCTTGTTGCATTTCTTAAAAAAGAGTAATAAAAA 1
```

RESULT 3
AL561661
LOCUS AL561661 995 bp mRNA linear EST 16-FEB-2001
DEFINITION AL561661 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL008YK21 5 prime, mRNA sequence.
ACCESSION AL561661
VERSION AL561661.1 GI:12909310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 995)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL008YK21"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 291 a 203 c 263 g 237 t 1 others
ORIGIN

Query Match 51.5%; Score 953.4; DB 9; Length 995;
Best Local Similarity 99.4%; Pred. No. 1.5e-218;
Matches 987; Conservative 1; Mismatches 2; Indels 3; Gaps 3;

```
Qy     24 TGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAG 83
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.
Db      1 TGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAG 60

Qy     84 GACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAA 143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAA 120

Qy    144 AGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGC 203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGC 180

Qy    204 AAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTC 263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTC 240

Qy    264 AGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGG 323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGG 300

Qy    324 TGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAA 383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAA 360

Qy    384 ATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTAC 443
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTAC 420

Qy    444 GGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGA 503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGA 480

Qy    504 ACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACA 563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACA 540

Qy    564 GGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACA 623
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACA 600

Qy    624 CCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAG 683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAG 660

Qy    684 CCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGT 743
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGT 720
```

```
Qy     744 GTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTC 803
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTC 780

Qy     804 AAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGT 863
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGT 840

Qy     864 CCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAA 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAA 900

Qy     924 CACTGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACAC 983
           |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||
Db     901 CACTGTAATTCTGTGGGATATGTCCTTGGGGAAA-CAGCAGCTAG-CTCGCTGTACACAC 958

Qy     984 AGACAAGGTCCAAACACTGCAGTTTCATCCATT 1016
           |||| |:||||||||||||||||||||||  ||
Db     959 AGAC-ARGTCCAAACACTGCAGTTTCATCATTT 990
```

RESULT 4
AL582812/c
LOCUS AL582812 1052 bp mRNA linear EST 16-FEB-2001
DEFINITION AL582812 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL008YK21 3 prime, mRNA sequence.
ACCESSION AL582812
VERSION AL582812.1 GI:12951167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1052)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES Location/Qualifiers
source 1. .1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL008YK21"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 279 a 237 c 205 g 303 t 28 others
ORIGIN

Query Match 51.3%; Score 950.6; DB 9; Length 1052;
Best Local Similarity 95.7%; Pred. No. 6.9e-218;
Matches 1004; Conservative 23; Mismatches 16; Indels 6; Gaps 5;

```
Qy     747 GGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAA 806
           |||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1047 GGACCTTKATATAGT-GACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAA 989

Qy     807 AAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCT 866
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
```

```
Db     988 AAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAA-GGCATACCGATGCTGTCCT 930

Qy     867 TGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACAC 926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 TGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACAC 870

Qy     927 TGTAATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGA 986
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     869 TGTAATTCTGTGGGATATGTCCTT-GGGAAACCAGCAGCTAGCCTCGCTGTACACACAGA 811

Qy     987 CAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATA 1046
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     810 CAAGGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATA 751

Qy    1047 TGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCG 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     750 TGATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCG 691

Qy    1107 ATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGC 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     690 ATTCAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGC 631

Qy    1167 CAGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTAC 1226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     630 CAGTACAGATGACGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTAC 571

Qy    1227 ACTTAATGCACACAATGATGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTG 1286
           |||||||||||||||||  ||| ||||||||||||||||||||||||||||||||||||
Db     570 ACTTAATGCACACAATGAAAAAAACTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTG 511

Qy    1287 TCTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCC 1346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     510 TCTCGTGACTGCTTCAGCTGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCC 451

Qy    1347 AAGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCC 1406
           |||||||||||||||||||||||||||||||||||||||||||| :|||||||||||||||
Db     450 AAGTCTAGTTCATTCTAGGGACATGAAAATGGGAGTTCTCTAMTGTTCTTCATGTTGCCC 391

Qy    1407 TGATTTGCCATTTATTTATGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATAT 1466
           |||||||||| |:||| |:|||||||||||||||||||||||||||:||||||||||
Db     390 TGATTTGCCAATWATTAAWGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGRTCTGGGATAT 331

Qy    1467 AAGCACAGTCTCTTCAGTAAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAG 1526
           ||||||| |||||:||||||||||:|||||||||||||||::| || |||:|||||
Db     330 AAGCACAMTCTCTTMAGTAAATGAAGMATTTGGAAGACGAGAGAVVCMTGATCTKGGGAG 271

Qy    1527 TGCAAGAAATTCATCTATTAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGA 1586
           |||||||| ||||||| |||||||::|||||||||||||||| :|||:||||:||||||
Db     270 GGCAARAAAATCATCWAATAGTGGCCSKTTTGGCAGCAGGAGCVVAGAWACACSCATGGA 211

Qy    1587 GTCTTAATGAAGATCATCTAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGC 1646
           |||||||||||||||||||||||| ||||||||||||||||::|||||||||||||||
Db     210 GTCTTAATGAAGATCATCTAATTTCCGGCTTACCTTAACTGGGAATKBTAAAAAGTTGGC 151

Qy    1647 CTAAAAATGTTCCATGCGTGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCT 1706
           ||||||:|||:|||||||||||:||| |||||||||||||:|||||||||||||||||||
Db     150 CTAAAARTGTKCCATGCGTGGCRGCAKCCATGCAGAGTGASTGAAACACAATTCATTTCT 91

Qy    1707 GACTGA-CATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGC 1765
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90 GACTGAGCATTCCTTTCTGCAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGC 31

Qy    1766 TCTTCA--GATGGATGGTTTGTAAGGCTC 1792
           ||||||   |||||||||||||||||||||
Db      30 TCTTCACGCATGGATGGTTTGTAAGGCTC 2
```

RESULT 5
AL515985
LOCUS AL515985 983 bp mRNA linear EST 13-FEB-2001
DEFINITION AL515985 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA001YJ04 5 prime, mRNA sequence.
ACCESSION AL515985

```
VERSION     AL515985.1  GI:12779478
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DA001YJ04"
                     /clone_lib="LTI_NFL011_NBC1"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library is not normalized, but is the control for
                     the normalized libraries. Library was constructed by Life
                     Technologies.  Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      284 a    203 c    258 g    233 t      5 others
ORIGIN
```

Query Match 50.6%; Score 938; DB 9; Length 983;
Best Local Similarity 99.4%; Pred. No. 7.4e-215;
Matches 934; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
Qy      19 AGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGAC 78
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGAC 60

Qy      79 TTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTG 138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTG 120

Qy     139 GCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCT 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCT 180

Qy     199 GAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGT 258
           ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db     181 GAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTKATGAAGAGGAGACAGGCAGT 240

Qy     259 CCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAG 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAG 300

Qy     319 GATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTA 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTA 360

Qy     379 GATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGT 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGT 420

Qy     439 CTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAA 498
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     421 CTTACGGTCTACGGGAGTTATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAA 480

Qy     499 TATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCT 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCT 540

Qy     559 GAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTAT 618
           ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db     541 GAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATHATCAAGAAGAAGACTCTTTTTAT 600

Qy     619 GTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGAT 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGAT 660

Qy     679 CCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATT 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATT 720

Qy     739 GAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAA 798
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAA 780

Qy     799 CTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGAT 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGAT 840

Qy     859 GCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCT 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCT 900

Qy     919 GACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAAAC 958
           ||||:||||||:|||||||||||||||||||||||||| |
Db     901 GACAMCACTGTWATTCTGTGGGATATGTCCTTGGGGAACC 940
```

```
RESULT 6
BM547947
LOCUS       BM547947                1073 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6531563 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5732716
            5', mRNA sequence.
ACCESSION   BM547947
VERSION     BM547947.1  GI:18782142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1073)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12735  row: e  column: 05
            High quality sequence start: 109
            High quality sequence stop: 739.
FEATURES             Location/Qualifiers
     source          1. .1073
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5732716"
                     /clone_lib="NIH_MGC_124"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
                     (destroyed); Site_2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
```

insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."

BASE COUNT 308 a 220 c 280 g 262 t 3 others
ORIGIN

Query Match 50.3%; Score 932; DB 13; Length 1073;
Best Local Similarity 96.4%; Pred. No. 2.1e-213;
Matches 1006; Conservative 0; Mismatches 32; Indels 6; Gaps 5;

```
Qy      1 GATCCCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 60
          | |||    |    |   ||  ||||||||||||||||||||||||||||||||||||||||||||
Db      6 GGTCCGGAATTCTCCGGATGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTA 65

Qy     61 GCGTGACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCC 120
          |||||||||||||||||||||||||||||||||||| ||||  ||| ||||||||  |||||
Db     66 GCGTGACACGCCCTTGACTTGAGGACCATGAACCTCAGCTCCCAAGTGACGTGTTTGGCC 125

Qy    121 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 TGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAA 185

Qy    181 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGAT 245

Qy    241 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 GAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGC 305

Qy    301 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAG 360
          ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db    306 CCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGATTTTTGATGATGATGAG 365

Qy    361 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 CTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTT 425

Qy    421 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACT 485

Qy    481 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486 CTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTT 545

Qy    541 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 ATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAA 605

Qy    601 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 GAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTG 665

Qy    661 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 GAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGA 725

Qy    721 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    726 AACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTC 785

Qy    781 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    786 TTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCA 845

Qy    841 GCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGG-AATAAGCTAATCAGAAATGT 899
          |||||| ||||||||||||||||||||||| ||||||| ||||||||||||||||||||
Db    846 GCAGAAAGGCATACCGATGCTGTCCTTGACCCTTCATGGNAATAAGCTAATCAGAAATGT 905

Qy    900 TTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTT-GGGGAAAC 958
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    906 NTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGGAAAC 965

Qy    959 CAGCAGCTAGCCT-CGCTGTACACACA-GACAAGGTCCAAACACTG--CAGTTTCATCCA 1014
          || |||||||||| |||||||||||||| ||||||||||||||||     |||||||||
Db    966 CANCAGCTAGCCTCCGCTGTACACACAGGACAAGGTCCAAACACTGGCAGTTTTCATCCA 1025

Qy   1015 TTTGAAGCACAGACTCTGATTTCT 1038
          ||| |||| |||||||||  || |
Db   1026 TTTTAAGCCCAGACTCTGGATTTT 1049
```

RESULT 7
AL549213
LOCUS AL549213 942 bp mRNA linear EST 16-FEB-2001
DEFINITION AL549213 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI049YG12 5 prime, mRNA sequence.
ACCESSION AL549213
VERSION AL549213.1 GI:12884971
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 942)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI049YG12"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 272 a 184 c 247 g 228 t 11 others
ORIGIN

Query Match 49.5%; Score 917; DB 9; Length 942;
Best Local Similarity 98.1%; Pred. No. 8.2e-210;
Matches 925; Conservative 11; Mismatches 6; Indels 1; Gaps 1;

```
Qy      5 CCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGT 64
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGT 60

Qy     65 GACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGG 124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGG 120

Qy    125 TCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAA 184
          |||||||||||||||||||||:|||||||||||||| |||||||||:||:||||||
Db    121 TCCGCTGCGGCGTGGCCAAAGAGAMACCAGACAAGGTAGAACTGAGTAAARAARAAGTAA 180

Qy    185 AACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAG 244
          |||||||||||||||||||||||||||||||||:||:||||||||||||||||||||:
Db    181 AACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAARAARAAGGTGGTGGCAGTGATGAAR 240

Qy    245 AGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAA 304
          ||||:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 AGGARAMAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAA 300

Qy    305 GAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGG 364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGG 360

Qy    365 CTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTG 424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTG 420

Qy    425 AATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGA 484
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    421 AATCTCTCTTGGGTCTTACGGTCTACGGGAGTTATGATCAAGATCCTTACGTTACTCTGA 480

Qy    485 AAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAG 544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAG 540

Qy    545 TTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAG 604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAG 600

Qy    605 AAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAAT 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAAT 660

Qy    665 GGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACA 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACA 720

Qy    725 TGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCA 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCA 780

Qy    785 CACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAG 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAG 840

Qy    845 AAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTTAG 904
          ||||||||| |||||||||| |: :| ||||||||||||||||||||||||||||||||
Db    841 AAGGGCATA-CGATGCTGTCTTKAMCTTTTCATGGAATAAGCTAATCAGAAATGTTTTAG 899

Qy    905 CAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTC 947
          |||||||||||||||||||||||| |||||||:|||||||||
Db    900 CAAGTGCATCAGCTGACAACACTGTTATTCTGTKGGATATGTC 942


RESULT 8
BQ228676
LOCUS       BQ228676                906 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7594492 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6061590
            5', mRNA sequence.
ACCESSION   BQ228676
VERSION     BQ228676.1  GI:20410076
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13331  row: 1  column: 07
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6061590"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      267 a    175 c    203 g    261 t
ORIGIN

  Query Match             48.9%;  Score 906;  DB 14;  Length 906;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-207;
  Matches  906;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    417 TCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGT 476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGT 60

Qy    477 TACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAA 536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAA 120

Qy    537 TCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAA 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAA 180

Qy    597 TCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAG 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAG 240

Qy    657 TGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGT 716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGT 300

Qy    717 AGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCC 776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCC 360

Qy    777 AGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTC 836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTC 420

Qy    837 CTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAA 896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGAATAAGCTAATCAGAAA 480

Qy    897 TGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAA 956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATATGTCCTTGGGGAA 540

Qy    957 ACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATT 1016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCATT 600

Qy   1017 TGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCG 1076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGCCG 660

Qy   1077 AAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTG 1136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACTTG 720

Qy   1137 GAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTT 1196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAATTT 780

Qy   1197 GGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGG 1256
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCTGG 840

Qy    1257 TCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGT 1316
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATACGT 900

Qy    1317 GAAGAT 1322
                ||||||
Db     901 GAAGAT 906


RESULT 9
AL576190/c
LOCUS       AL576190                902 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL576190 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI072YE01 3
            prime, mRNA sequence.
ACCESSION   AL576190
VERSION     AL576190.1  GI:12938088
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI072YE01"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      267 a    205 c    180 g    246 t      4 others
ORIGIN

  Query Match             47.5%;  Score 880;  DB 9;  Length 902;
  Best Local Similarity   99.2%;  Pred. No. 6.5e-201;
  Matches  891;  Conservative    3;  Mismatches    3;  Indels    1;  Gaps    1;

Qy     885 GCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATAT 944
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     898 GYTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCTGTGGGATAT 839

Qy     945 GTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCA 1004
           |||:|| ||||||||||||||||||||||:|||||||||||||  ||||||||||||
Db     838 GTCYTT-GGGAAACCAGCAGCTAGCCTCGCTGWACACACAGACAAGTACCAAACACTGCA 780

Qy    1005 GTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTT 1064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     779 GTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTT 720

Qy    1065 GTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGA 1124
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     719 GTATAACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGA 660

Qy    1125 GAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTT 1184
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     659 GAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTT 600

Qy    1185 TGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGA 1244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     599 TGTATATAATTTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGA 540

Qy    1245 TGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGC 1304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 TGAAATCTCTGGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGC 480

Qy    1305 TGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAG 1364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 TGACAAATACGTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAG 420

Qy    1365 GGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTA 1424
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 GGACATGAAAATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTA 360

Qy    1425 TGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGT 1484
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 TGCCTTTGGAGGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGT 300

Qy    1485 AAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTAT 1544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 AAATGAAGCATTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTAT 240

Qy    1545 TAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATC 1604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 TAGTGGCCCTTTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATC 180

Qy    1605 TAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCG 1664
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 TAATTTCCTGCTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCG 120

Qy    1665 TGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTG 1724
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     119 TGGCAGCAACCATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTG 60

Qy    1725 CAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTT 1782
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59 CAACTGCGGTGGCACCACAAATATCCGGTCTTTGTGCTTGCTCTTCAGATGGATGGTT 2


RESULT 10
BM552364
LOCUS       BM552364                1007 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6543589 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5549650
            5', mRNA sequence.
ACCESSION   BM552364
VERSION     BM552364.1  GI:18790164
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12260  row: i  column: 11
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .1007
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5549650"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      287 a    212 c    272 g    234 t      2 others
ORIGIN

  Query Match             46.8%;  Score 868;  DB 13;  Length 1007;
  Best Local Similarity   98.6%;  Pred. No. 5.1e-198;
  Matches  885;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       6 CTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTG 65
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGAGCGTGTGGCAGCAGTGCGGTCGTGGTCCCTCCCTATGCAGCCTGGTTTCTAGCGTG 60

Qy      66 ACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGT 125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACACGCCCTTGACTTGAGGACCATGAACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGT 120

Qy     126 CCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAA 185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAA 180

Qy     186 ACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGA 245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGA 240

Qy     246 GGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAG 305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAGTGCACGCACCCAGGCACGCCCAAG 300

Qy     306 AGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGC 365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAGGACGCTTGATGATGATGAGCTGGC 360

Qy     366 TGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGA 425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGAGTACGACTTAGATAAATATGATGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGA 420

Qy     426 ATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAA 485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGATCAAGATCCTTACGTTACTCTGAA 480

Qy     486 AGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGT 545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGATACAGAACAATATGAACGTGAAGATTTCTTGATTAAGCCCAGTGATAATCTTATAGT 540

Qy     546 TTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGA 605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGA 600

Qy     606 AGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATG 665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGACTCTTTTTATGTACACCATGATATACTCTTGTCTGCATATCCTCTGAGTGTGGAATG 660

Qy     666 GCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTAGGAAACAT 725
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     661 GCTGAATTTTGATCCTAGCCCAGATGATTCTACTGGAAATTACATTGCTGTANGAAACAT 720

Qy     726 GACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCAC 785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GACCCCTGTTATTGAAGTGTGGGACCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCAC 780

Qy     786 ACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGA 845
           |||||||||||||||||||||||||||||| |||||||| ||  |||||||||||||||||
Db     781 ACTCGGAAGTAAACTTTCAAAAAAGAAGAANAAGAAAGGGAAAGAGAGTTCCTCAGCAGA 840

Qy     846 AGGGCATACCGATGCTGTCC-TTGACCTTTCATGGAATAAGCTAATCAGAAATGTTTT 902
           | |||||||||||||||| |||||||||||||||||||||||||||||   ||  |||
Db     841 AAGGCATACCGATGCTGTCCTTTGACCTTTCATGGAATAAGCTAATCCAGAAATGTTT 898


RESULT 11
BQ948509
LOCUS       BQ948509                894 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8795115 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:6427975
            5', mRNA sequence.
ACCESSION   BQ948509
VERSION     BQ948509.1  GI:22363987
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2613  row: n  column: 08
            High quality sequence start: 14
            High quality sequence stop: 670.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6427975"
                     /clone_lib="NIH_MGC_101"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      264 a    167 c    210 g    252 t      1 others
ORIGIN

  Query Match             46.7%;  Score 865;  DB 14;  Length 894;
  Best Local Similarity   98.7%;  Pred. No. 2.6e-197;
  Matches  882;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy     330 AGAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGA 389
           |||||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGAGGATGACAGGATTCCTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGA 60

Qy     390 TGAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTA 449
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db      61 TGAGGAAGGTGACCCAGATGTTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTA 120

Qy     450 CGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGA 509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGA 180

Qy     510 AGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCA 569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCA 240

Qy     570 GTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGA 629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 GTGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGA 300

Qy    630 TATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGA 689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGA 360

Qy    690 TGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGA 749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGA 420

Qy    750 CCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAA 809
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAA 480

Qy    810 GAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGA 869
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGA 540

Qy    870 CCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGT 929
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGT 600

Qy    930 AATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAA 989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATTCTGTGGGATATGTCCTTGGGGAAACCAGCAGCTAGCCTCGCTGTACACACAGACAA 660

Qy    990 GGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGA 1049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGTCCAAACACTGCAGTTTCATCCATTTGAAGCACAGACTCTGATTTCTGGCTCATATGA 720

Qy   1050 TAAGTCAGTGGCTTTGTATGACTGCCGAAGTCCAGATGAAAGCCATCGAATGTGGCGATT 1109
          |||||||||||||||||||||||| |   ||||||||||||||||||||||||||||||
Db    721 TAAGTCAGTGGCTTTGTATGACTGCCNGAAGTCAGATGAAAGCCATCGAATGTGGCGATT 780

Qy   1110 CAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAG 1169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGTGGGCAGATAGAGAGAGTGACTTGGAATCACTTTTCACCTTGTCATTTCTTGGCCAG 840

Qy   1170 TACAGATGA-CGGCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCAATTT 1222
          ||||||||| || |||||||||||||||||||||||||||||||||| ||||
Db    841 TACAGATGACCGCCTTTGTATATAATTTGGATGCACGTTCAGATAAGCCCATTT 894


RESULT 12
BQ224661
LOCUS       BQ224661                869 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7569692 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043941
            5', mRNA sequence.
ACCESSION   BQ224661
VERSION     BQ224661.1  GI:20406061
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13285  row: l  column: 22
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1. .869
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6043941"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      257 a    172 c    233 g    207 t
ORIGIN

  Query Match             45.7%;  Score 847.4;  DB 14;  Length 869;
  Best Local Similarity   99.2%;  Pred. No. 4.5e-193;
  Matches  862;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy     37 CCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGC 96
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCTCCCTATGCAGCCTGGTTTCTAGCGTGACACGCCCTTGACTTGAGGACCATGAACCGC 60

Qy     97 AGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGAC 156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGAC 120

Qy    157 AAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTG 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTG 180

Qy    217 CAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATG 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATG 240

Qy    277 CAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGAT 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGAT 300

Qy    337 GACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAA 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAA 360

Qy    397 GGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGT 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGT 420

Qy    457 AATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTC 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTC 480

Qy    517 TTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAAT 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAAT 540

Qy    577 TTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTC 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTC 600

Qy    637 TTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCT 696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCT 660

Qy    697 ACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGAT 756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGAT 720

Qy    757 ATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAA 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAA 780

Qy    817 AAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATA-CCGATGCTGTCCTTGACCTTTC 875
          |||||||||||||||||||||||||||||||||||| |||||||||||||||   |||||
Db    781 AAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCCGATGCTGTCCTTTGACTTTC 840
```

```
Qy    876 ATGGAATAAGCTAATCAGAAATGTTTTAG 904
          ||||  |  ||||||||||||||||||||||
Db    841 ATGGGAATAGCTAATCAGAAATGTTTTAG 869


RESULT 13
BQ643631
LOCUS       BQ643631                953 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8342961 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268799
            5', mRNA sequence.
ACCESSION   BQ643631
VERSION     BQ643631.1  GI:21767803
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 953)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2441  row: e  column: 24
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .953
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6268799"
                     /clone_lib="NIH_MGC_100"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      272 a    187 c    224 g    270 t
ORIGIN

  Query Match             45.5%;  Score 842.4;  DB 14;  Length 953;
  Best Local Similarity   98.3%;  Pred. No. 7.2e-192;
  Matches  873;  Conservative    0;  Mismatches   11;  Indels    4;  Gaps    2;

Qy    955 AAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCA 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAACCAGCAGCTAGCCTCGCTGTACACACAGACAAGGTCCAAACACTGCAGTTTCATCCA 60

Qy   1015 TTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTGAAGCACAGACTCTGATTTCTGGCTCATATGATAAGTCAGTGGCTTTGTATGACTGC 120

Qy   1075 CGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACT 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGAAGTCCAGATGAAAGCCATCGAATGTGGCGATTCAGTGGGCAGATAGAGAGAGTGACT 180

Qy   1135 TGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAAT 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGAATCACTTTTCACCTTGTCATTTCTTGGCCAGTACAGATGACGGCTTTGTATATAAT 240

Qy   1195 TTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCT 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTGGATGCACGTTCAGATAAGCCAATTTTTACACTTAATGCACACAATGATGAAATCTCT 300

Qy   1255 GGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATAC 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGTCTTGATCTTAGCAGTCAAATCAAGGGCTGTCTCGTGACTGCTTCAGCTGACAAATAC 360

Qy   1315 GTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAA 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGAAGATCTGGGACATCTTAGGAGATAGGCCAAGTCTAGTTCATTCTAGGGACATGAAA 420

Qy   1375 ATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGA 1434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGAGTTCTCTTCTGTTCTTCATGTTGCCCTGATTTGCCATTTATTTATGCCTTTGGA 480

Qy   1435 GGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCA 1494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGTCAAAAAGAAGGGCTTCGGGTCTGGGATATAAGCACAGTCTCTTCAGTAAATGAAGCA 540

Qy   1495 TTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCT 1554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTGGAAGACGAGAGAGGCTTGTTCTTGGGAGTGCAAGAAATTCATCTATTAGTGGCCCT 600

Qy   1555 TTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTG 1614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTTGGCAGCAGGAGCTCAGATACACCCATGGAGTCTTAATGAAGATCATCTAATTTCCTG 660

Qy   1615 CTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAAC 1674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTTACCTTAACTGGGAATTTTAAAAAGTTGGCCTAAAAATGTTCCATGCGTGGCAGCAAC 720

Qy   1675 CATGCAGAGTGACTGAAACACAATTCATTTCTGACTGACATTCCTTTCTGCAACTGCGGT 1734
          |||||||||||||||||||||||| ||||||| ||||||||||||||  ||||||||||||
Db    721 CATGCAGAGTGACTGAAACACAATTCCTTTCTGAATGACATTCCTTTCTGGCACTGCGGT 780

Qy   1735 GGCACCACAAATATCCGGTCTTTGTGCTT-GCTCTTCAGATGGATGGTTTGTAAGGCTCT 1793
          ||||||||||||||||||| |||||||||| |||||||||||||||||||||||||||||
Db    781 GGCACCACAAATATCCGGGCTTTGTGCTTGGCTCTTCAGATGGATGGTTTGTAAGGCTCT 840

Qy   1794 TGTTGCATTTCTTAAAAAAGAGTAATAAAAAGGATTTTTAAAAAGTAA 1841
          ||||||||||||||||   ||      ||||||||||||||||||||
Db    841 TGTTGCATTTCTTAAAA---AGAGTATTAAAGGATTTTTAAAAAGTAA 885


RESULT 14
BQ646683
LOCUS       BQ646683                941 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8342479 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267952
            5', mRNA sequence.
ACCESSION   BQ646683
VERSION     BQ646683.1  GI:21770855
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2439  row: b  column: 17
            High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .941
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6267952"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 280 a 189 c 248 g 223 t 1 others
ORIGIN

Query Match 44.5%; Score 825.4; DB 14; Length 941;
Best Local Similarity 99.0%; Pred. No. 8.9e-188;
Matches 851; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

```
Qy    102 CCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGT 161
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCTGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACACCAGACAAGGT 60

Qy    162 AGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGA 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAGAAATTGCAAGA 120

Qy    222 AGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAG 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGATGGCATGCAGAG 180

Qy    282 TGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAG 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCAGAGGATGACAG 240

Qy    342 GACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGA 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGATGAGGAAGGTGA 300

Qy    402 CCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGA 461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTACGGGAGTAATGA 360

Qy    462 TCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGAT 521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAAGATTTCTTGAT 420

Qy    522 TAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGA 581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAGTGCAATTTAGA 480

Qy    582 GGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTC 641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGATATACTCTTGTC 540

Qy    642 TGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGG 701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGATGATTCTACTGG 600

Qy    702 AAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGT 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGACCTTGATATAGT 660

Qy    762 GGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAA 821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAGAAGAAAAAGAA 720

Qy    822 AGGAAAGAA-GAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGA 880
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGGAAAGAAGGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGACCTTTCATGGA 780
```

```
Qy    881 ATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCAGCTGACAACACTGTAATTCT-GTGG 939
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||| | ||
Db    781 ATAAGCTAATCAGAAATGTTTTAGCAAGTGCATCANCTGACAACACTGTAATTCTGGGGG 840

Qy    940 GATATGTCCTTGGGGAAACC 959
          ||||||||| | ||| || |
Db    841 GATATGTCCCTTGGGGAAAC 860
```

RESULT 15
BQ651718
LOCUS BQ651718 965 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8303996 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269343 5', mRNA sequence.
ACCESSION BQ651718
VERSION BQ651718.1 GI:21775890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 965)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2442 row: l column: 16
High quality sequence stop: 645.
FEATURES Location/Qualifiers
source 1. .965
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6269343"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 286 a 202 c 257 g 219 t 1 others
ORIGIN

Query Match 44.4%; Score 823.2; DB 14; Length 965;
Best Local Similarity 98.2%; Pred. No. 3e-187;
Matches 842; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```
Qy     91 AACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACA 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACCGCAGCCGCCAGGTGACGTGCGTGGCCTGGGTCCGCTGCGGCGTGGCCAAAGAGACA 60

Qy    151 CCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAG 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAGACAAGGTAGAGCTGAGTAAAGAAGAAGTAAAACGCCTCATTGCTGAGGCAAAGGAG 120

Qy    211 AAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGAT 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAATTGCAAGAAGAAGGTGGTGGCAGTGATGAAGAGGAGACAGGCAGTCCTTCAGAAGAT 180
```

```
Qy    271 GGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCA 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCATGCAGAGTGCACGCACCCAGGCACGCCCAAGAGAGCCCCTGGAGGATGGTGACCCA 240

Qy    331 GAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGAT 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGATGACAGGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTAGATAAATATGAT 300

Qy    391 GAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTAC 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGGAAGGTGACCCAGATGCTGAGACTCTTGGTGAATCTCTCTTGGGTCTTACGGTCTAC 360

Qy    451 GGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAA 510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGGAGTAATGATCAAGATCCTTACGTTACTCTGAAAGATACAGAACAATATGAACGTGAA 420

Qy    511 GATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAG 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GATTTCTTGATTAAGCCCAGTGATAATCTTATAGTTTGTGGCCGAGCTGAACAGGACCAG 480

Qy    571 TGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGAT 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCAATTTAGAGGTGCATGTTTATAATCAAGAAGAAGACTCTTTTTATGTACACCATGAT 540

Qy    631 ATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGAT 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATACTCTTGTCTGCATATCCTCTGAGTGTGGAATGGCTGAATTTTGATCCTAGCCCAGAT 600

Qy    691 GATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGAC 750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GATTCTACTGGAAATTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGTGGGAC 660

Qy    751 CTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCAAAAAAG 810
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    661 CTTGATATAGTGGACTCTTTAGAGCCAGTCTTCACACTCGGAAGTAAACTTTCNAAAAAG 720

Qy    811 AAGAAAAAGAAAGGAAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGAC 870
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AAGAAAAAGAAAGGGAAGAAGAGTTCCTCAGCAGAAGGGCATACCGATGCTGTCCTTGAC 780

Qy    871 CTTTCATGGAATAAGCTAATCAGAAATGTTTTAGCAAGT-GCATCAGCTGACAACACTGT 929
          | |||||||||  ||||||||||||||||||||||| ||| ||||||||||||||||||||
Db    781 CCTTCATGGAAATAGCTAATCAGAAATGTTTTAGCCAGTGGCATCAGCTGACAACACTGT 840

Qy    930 AATTCTGTGGGATATGT 946
          || |    |||  || |
Db    841 AAATTCTGGGGGGATAT 857
```

Search completed: March 19, 2003, 18:16:54
Job time : 9961 secs